MULTIVARIATE MULTISITE STATISTICAL DOWNSCALING OF ATMOSPHERE-OCEAN GENERAL CIRCULATION MODEL OUTPUTS OVER THE CANADIAN PRAIRIE PROVINCES

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In School of Environment and Sustainability

University of Saskatchewan

Saskatoon

By

ZILEFAC ELVIS ASONG

PERMISSION TO USE STATEMENT

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Executive Director

School of Environment and Sustainability

University of Saskatchewan

Room 323, Kirk Hall

117 Science Place

Saskatoon, SK S7N 5C8, Canada

Phone: (306) 966-1985

Fax: (306) 966-2298

Email: sens.info@usask.ca

ABSTRACT

Atmosphere-Ocean General Circulation Models (AOGCMs) are the primary tool for modelling global climate change in the future. However, their coarse spatial resolution does not permit direct application for local scale impact studies. Therefore, either dynamical or statistical downscaling techniques are used for translating AOGCM outputs to local scale climatic variables.

The main goal of this study was to improve our understanding of the historical and future climate change at local-scale in the Canadian Prairie Provinces (CPPs) of Alberta, Saskatchewan and Manitoba, comprising 47 diverse watersheds. Given the vast nature of the study area and paucity of recorded data, a novel approach for identifying homogeneous regions for regionalization of precipitation characteristics for the CPPs was proposed. This approach incorporated information about predictors — large-scale atmospheric covariates from the National Center for Environmental Prediction (NCEP) Reanalysis-I, teleconnection indices and geographical site attributes that impact spatial patterns of precipitation in order to delineate homogeneous precipitation regions using a combination of multivariate approaches. This resulted in the delimitation of five homogeneous climatic regions which were validated independently for homogeneity using statistics computed from observations recorded at 120 stations across the CPPs.

For multisite multivariate statistical downscaling, an approach based on the Generalized Linear Model (GLM) framework was developed to downscale daily observations of precipitation and minimum and maximum temperatures from 120 sites located across the CPPs. First, the aforementioned predictors and observed daily precipitation and temperature records were used to calibrate GLMs for the 1971–2000 period. Then the calibrated GLMs were used to generate daily sequences of precipitation and temperatures for the 1962–2005 historical (conditioned on

NCEP predictors), and future period (2006–2100) using outputs from six CMIP5 (Coupled Model Intercomparison Project Phase-5) AOGCMs corresponding to Representative Concentration Pathway (RCP): RCP2.6, RCP4.5, and RCP8.5 scenarios. The results indicated that the fitted GLMs were able to capture spatiotemporal characteristics of observed climatic fields. According to the downscaled future climate, mean precipitation is projected to increase in summer and decrease in winter while minimum temperature is expected to warm faster than the maximum temperature. Climate extremes are projected to intensify with increased radiative forcing.

ACKNOWLEDGMENTS

I would like to thank my supervisor Dr. Naveed Khaliq and co-supervisor Prof. Howard Wheater for their invaluable support and advice throughout my PhD research, as well as the members of my advisory committee: Dr. Yanping Li, Dr. Andrew Ireson, Dr. Karl-Erich Lindenschmidt, and Prof. Amin Elshorbagy. The valuable comments and suggestions provided by them were very helpful to me in successfully writing the conference and journal manuscripts on which this thesis is based.

The financial support from the Canada Excellence Research Chair in Water Security, School of Environment and Sustainability and the Government of Saskatchewan is gratefully acknowledged. Thanks are due to Eva Mekis from Environment Canada for providing access to observed precipitation and temperature data used in this study. I would like to thank my friends at the Global Institute for Water Security who helped in many different ways during my PhD research.

Finally, an honorable mention goes to my family and friends. I deeply owe my father ―Ndi Nkemasongafac Eric, my mother―Nkemasongafac Justine, and all my brothers and sisters. Without the support of the above mentioned people, this project would not have been a success.

DEDICATION

This thesis is dedicated to my wife, Acha Marvel Emefeut, our kids and to God Almighty

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACF	Auto-Correlation Function
AO	Artic Oscillation
AOGCM	Atmosphere-Ocean General Climate Model
AR5	Assessment Report 5
CA	Cluster Analysis
CAPE	Convective Available Potential Energy
CCA	Canonical Correlation Analysis
CIN	Convective Inhibition
CMIP5	Coupled Model Intercomparison Project 5
CPPs	Canadian Prairie Provinces
DJF	December January February
ENSO	El Nino Southern Oscillation
EOF	Empirical Orthogonal Function
ESG	Earth System Grid
FAO	Food and Agricultural Organization
FCM	Fuzzy C-Means
GCM	General Climate Model
GEV	Generalized Extreme Value
GHG	Green House Gas
GLM	Generalized Linear Model
GLO	Generalized Logistic
GNO	Generalized Normal
GPA	Generalized Pareto

IPCC	Inter-governmental Panel on Climate Change
JJA	June July August
LARS-WG	Long Ashton Research Station Weather
L-CV	L-Coefficient of Variation
MAM	March April May
MME	Multi-Model Ensemble
MK	Mann Kendall
NAO	North Atlantic Oscillation
NCEP/NCAR	National Center for Environmental Prediction/ National Center for Atmospheric Research
NP	North Pacific Oscillation
PC	Principal Component
PCA	Principal Component Analysis
PCMDI	Program for Climate Model Diagnosis and Intercomparison
PDO	Pacific Decadal Oscillation
PE3	Pearson Type III
PNA	Pacific North American Pattern
PP	Perfect Prognosis
PW	Proportion of Wet days
Q-Q	Quantile-Quantile
RCM	Regional Climate Model
RCP	Representative Concentration Pathway
RFA	Regional Frequency Analysis

SD	statistical downscaling
SON	September October November
UK	United Kingdom
Wak	Wakeby
WG	Weather Generator
WGEN	Weather GENerator

CHAPTER 1
INTRODUCTION

1.1 Climate Change Impacts on Water Resources

Water resources are indispensable for society, ecosystems and all forms of life. To sustain health, humanity depends on a reliable and accessible clean supply of drinking water. Water is valuable not only for domestic uses, but also for its role in maintaining aquatic ecosystems and other socio-environmental sectors, including agriculture, energy production, transportation, tourism and other industrial uses (Young, 2005). The majority of these water uses require freshwater which is quite limited compared to, for example, saltwater. In terms of the distribution of renewable water resources, there is extreme variability from country to country. This pattern is largely a function of the patchwork of climates and physiographic structures (FAO, 2003). With rising population growth, the global and regional demand for freshwater is projected to likely increase in the future especially under climate change conditions (Vörösmarty et al., 2000; Hejazi et al., 2013).

The variations in the distribution of surface and groundwater resources are governed largely by the spatial and temporal changes in the climatic variables such as precipitation, temperature, humidity and evaporation (Dingman, 2008). In more specific terms, the effective depth of precipitation over a given catchment plays a major role in its water budget while hydro-climatic variables such as evaporation, temperature, wind speed, atmospheric water vapor content determine some of the water losses from a catchment (Viessman and Lewis, 2012). Apart from water quantity, an increase in the temperature of water can alter the rate of chemical reactions and hence the water quality (Murdoch et al., 2000). Likewise, heavy precipitation events can increase the nutrient and sediment loads entrained by the rivers and hence affect the quality of water (Kundzewicz et al., 2007).

Findings from the Intergovernmental Panel on Climate Change (IPCC, 2013) Fifth Assessment Report (AR5) indicate that climate change due to increasing greenhouse gas (GHG) concentrations in the atmosphere continues to be a major concern. Climate change will have a range of impacts on water resources at different spatial and temporal time scales due to changes in precipitation (e.g., global average annual precipitation through the end of the 21st century is expected to increase, albeit changes in the amount and intensity of precipitation will likely vary by region) and temperature (e.g., AR5 projections estimate a likely 0.3°C to 0.7°C change in global mean surface temperature for the period 2016–2035 relative to 1986–2005) patterns which will lead to intensification of the hydrological cycle (Stocker et al., 2013). Consequently, changes in the spatiotemporal patterns of precipitation, sea level rise, increased snow and ice melt rates, changes in the characteristics of climate extremes, rise in atmospheric water vapor content, increase in evaporation, changes in the variability of runoff patterns, and erosion and sediment transport are anticipated. The potential impacts of a warming climate on water availability in snow-dominated regions continue to be a serious concern. Barnett et al. (2005) concluded that in a warmer world, less winter precipitation will probably fall as snow and the melting of winter snow may occur earlier in spring. A shift in peak river runoff to winter and early spring, away from summer and autumn when demand is highest is expected.

Given that climate change is very important for water resources decision support mechanisms in a river basin, information about projected changes in various climatic variables will be crucial for sustainable management of water resources. Canada in particular is one of the many countries which are affected by climate change. Zhang et al. (2000) performed an analysis of temperature and precipitation trends in Canada during the 20th century using station data. During the period 1900–1998, they found annual mean temperature to have increased between 0.5 and 1.5^{o}C in the southern regions with warming in minimum temperature greater than

maximum temperature in the first half of the century. Regionally, the greatest warming occurred in western Canada with statistically significant increases in spring and summer seasons. Similarly, across Canada, precipitation increased by 5 to 35%, with significant negative trends in the southern regions during winter. Vincent et al. (2015) found the ratio of snowfall to total precipitation to have decreased, with significant negative trends occurring mostly in southern Canada during spring. A review of recent climatic and cryospheric changes in this region is provided by DeBeer et al. (2015).

Throughout the Canadian Prairie Provinces (CPPs) of Alberta, Saskatchewan and Manitoba, there is increasing demand for freshwater resources due to population growth, urbanization, agricultural and industrial expansion. In contrast, the availability (over space and time) of future water resources is uncertain probably due to climate change, as decline in streamflows is expected to reduce water availability for agriculture, industrial and domestic use, as well as hydroelectric power generation. In addition, earlier spring runoff, reduced snow accumulations and extents, and decreased glacial melt contributions are some of the likely projected impacts of climate change in mountainous regions (Demuth and Pietroniro, 2003; Sauchyn and Kulshreshtha, 2008).

With over 37 million hectares of land under cultivation — 80% of Canada's farmland, the Prairies are Canada's agricultural powerhouse. The South Saskatchewan River Basin provides a distinct illustration of the fact that while irrigation expansion in the Prairies region may be highly desirable to improve agricultural productivity and protect food security against current and future climate variability, it cannot be considered in isolation from a set of competing pressures for alternative water uses, including the needs for energy production and ecosystem functioning (Council of Canadian Academies, 2013). The expected increases in the magnitude and frequency of drought conditions and the reduction in snowpack, which replenishes both the soil moisture

and groundwater storage during spring melt and produces the dominant component of streamflow in the Rocky Mountains region, will have relevant implications for irrigated agriculture (Nazemi et al., 2012; Wheater and Gober, 2013). As a result of these changing climatic conditions, demand for irrigation water is likely to increase at a time when critical ecosystem flows are at their minimum.

Furthermore, it is likely that changes in the seasonal and extreme precipitation characteristics will have various implications for coping with basin-wide water management-related problems in this region under an uncertain future (Khaliq et al., 2014). In the CPPs, about 75% of the population (~5.5 million people) live in urban areas. The population is expected to grow by 30% in the next 25 years, with 40% of that growth taking place in Alberta (Natural Resources Canada, 2010). Increased urbanisation may lead to intensified urban "heat island" effect, with an increase in paved surfaces and the need for improved flood forecasting, design flood estimation, flood protection structures and drainage networks to counter flash floods. On the contrary, a lack of precipitation will probably lead to pressures on already declining water resources and food supply, and effects can be severe, depending on the resilience of the local society and population in the face of a changing climate. Therefore, understanding these changes is critical to the solution of emerging and challenging environmental issues such as (1) the assessments of regional impacts on the sustainability of water resources under a changing climate, (2) the predictability of wet and dry extremes and their impacts on human wellbeing, and (3) the impact on catchment scale hydro-climatology due to climate change and land use, and the feedbacks of shifting hydrological regimes on the regional climate.

Atmosphere-Ocean General Circulation Models (AOGCMs) are the best tools available for the prediction of the impacts of natural and anthropogenic modifications of the atmosphere on the global climate, hundreds of years into the future. However, the horizontal resolution of these

models are often on the order of several hundred kilometers and hence cannot resolve the station scale climate features which are important in terms of water resources assessments in a watershed. In other words, although AOGCMs outputs can reasonably represent climatic features at continental and global scales, they remain relatively coarse and unable to resolve significant sub-grid scale features such as topography, clouds and land use (Johns et al., 1997). Apart from their coarse resolution, the uncertain reliability of their output on time scales of months or less, especially for variables pertaining directly to the hydrologic cycle necessitate bias correction techniques (Carter et al., 1994). Therefore, outputs from these models cannot reliably be applied directly in many climate change impacts assessment studies at the local and catchment scales. Thus, there is a need for translating this coarse information provided by AOGCMs to station scale hydro-climatic variables such as precipitation, temperature, and evapotranspiration. For this purpose, downscaling methods (i.e., statistical and dynamical) have often been widely utilized (Fowler et al., 2007).

The concept of downscaling using statistical and dynamic methods is graphically illustrated in Figure 1-1. As presented in this Figure for dynamical downscaling, the relationships between the coarse resolution AOGCM outputs and the relatively fine resolution outputs of the Regional Climate Model (RCM) are established by considering the atmospheric physics. On the contrary, statistical downscaling is based on the formulation of statistical relationships between the coarse resolution AOGCM outputs and the climatic variables at the points of interest (e.g., gauging stations) in the catchment. Further details on downscaling techniques are provided in Chapters 3 and 4 of this dissertation.

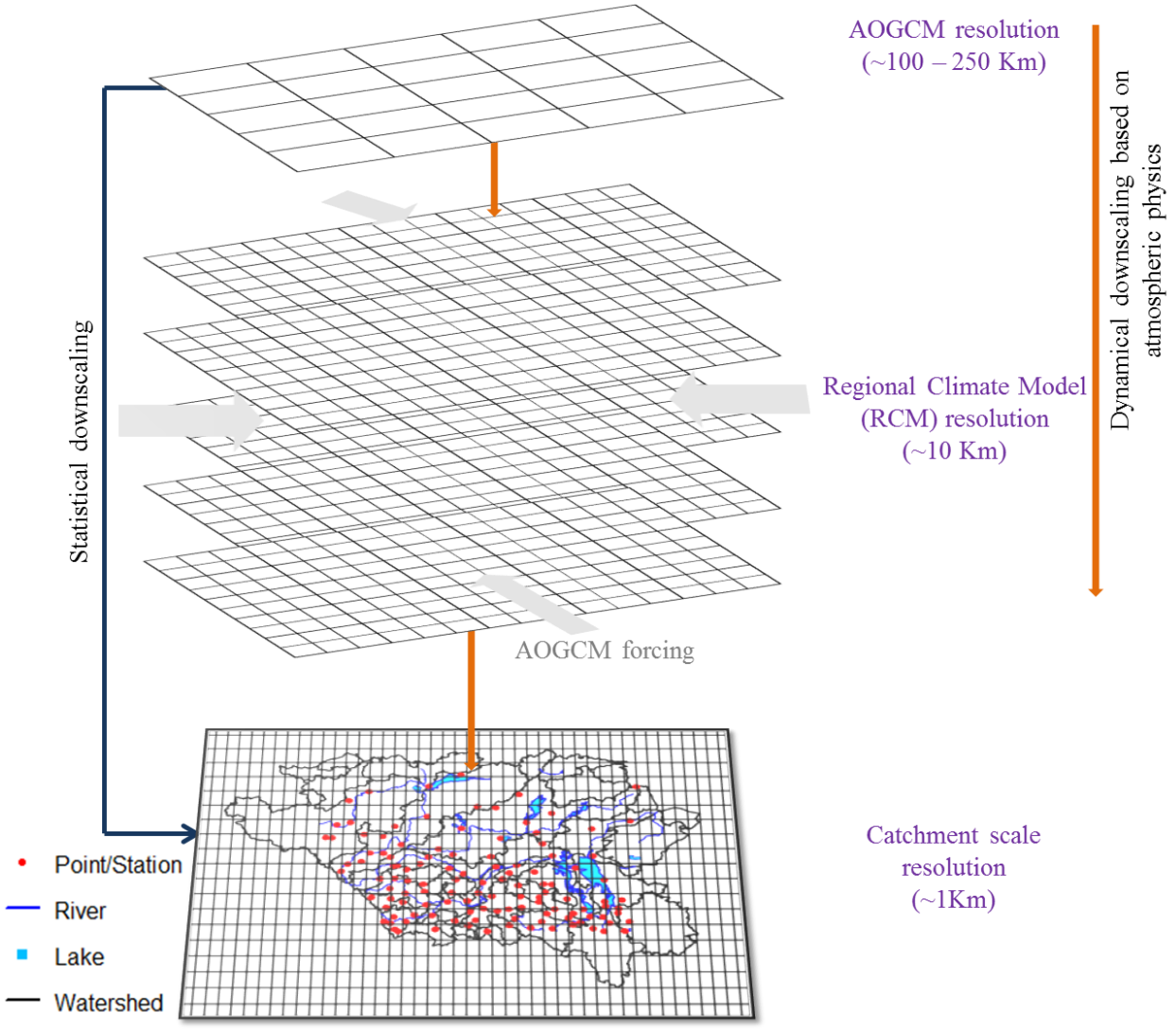

Figure 1-1: An illustration of statistical and dynamical downscaling concepts (adapted from http://epscorspo.nevada.edu/nsf/climate1/climate10.html)

1.2 Project Goal and Research Objectives

The main goal of this study was to improve our understanding of historical climate and future climate change at local, catchment and regional scales in the CPPs. To this end, a multisite multivariate statistical downscaling approach based on the Generalized Linear Model (GLM) framework was developed to downscale coarse AOGCM outputs to station scale climatic variables such as daily precipitation, and minimum and maximum temperatures. Within this overall goal, specific objectives were to:

(1) Identify large scale climate drivers of local scale climate anomalies and climate change in this region of Canada. Here, the aim was to explore and document the utility of large scale atmospheric circulation patterns in delimiting homogeneous precipitation climates through multivariate approaches and to evaluate independently the homogeneity of these regions using statistics obtained from station-based observed precipitation data at different time scales.

(2) Investigate the suitability of GLMs for multisite multivariate modelling of precipitation and temperature fields in the CPPs region comprising 47 diverse watersheds, with significant regional inhomogeneity and a paucity of ground-based observations.

(3) Use the developed GLM framework to downscale Coupled Model Intercomparison Project Phase 5 (CMIP5) AOGCM outputs corresponding to three Representative Concentration Pathway (RCP) scenarios to station scale climatic variables. This would provide plausible future scenarios of changes in characteristics of temperature and precipitation at a scale relevant for impact assessments and policymaking in the CPPs.

1.3 Study Area

The aim of the study and specific objectives outlined in Section 1.2 were demonstrated through a case study of the CPPs of Alberta, Saskatchewan and Manitoba. The location of the study region (that has a total surface area of 1,960,681 km^2) is shown in Figure 1-2. The ecosystems of this region depend heavily on precipitation and its timing (Hogg et al., 2000). Apart from the moderating effects due to regional changes in topography, atmospheric circulation also controls precipitation patterns (Borchert, 1950). Annual average precipitation is approximately 454 mm, rather less than the Canada-wide average of 535 mm (Phillip, 1990). The major inflows to the Saskatchewan River Basin, the largest river system in the region,

originate from the Rocky Mountains (Wheater and Gober, 2013). Characterized by a highly variable hydro-climate and diminishing water resources (Bonsal et al., 2012), southern parts of this region support a vibrant agro-based economy that was hard-hit by the most severe and prolonged droughts of 1988 and 1999–2005, as well as severe floods of 2011, 2013 and 2014.

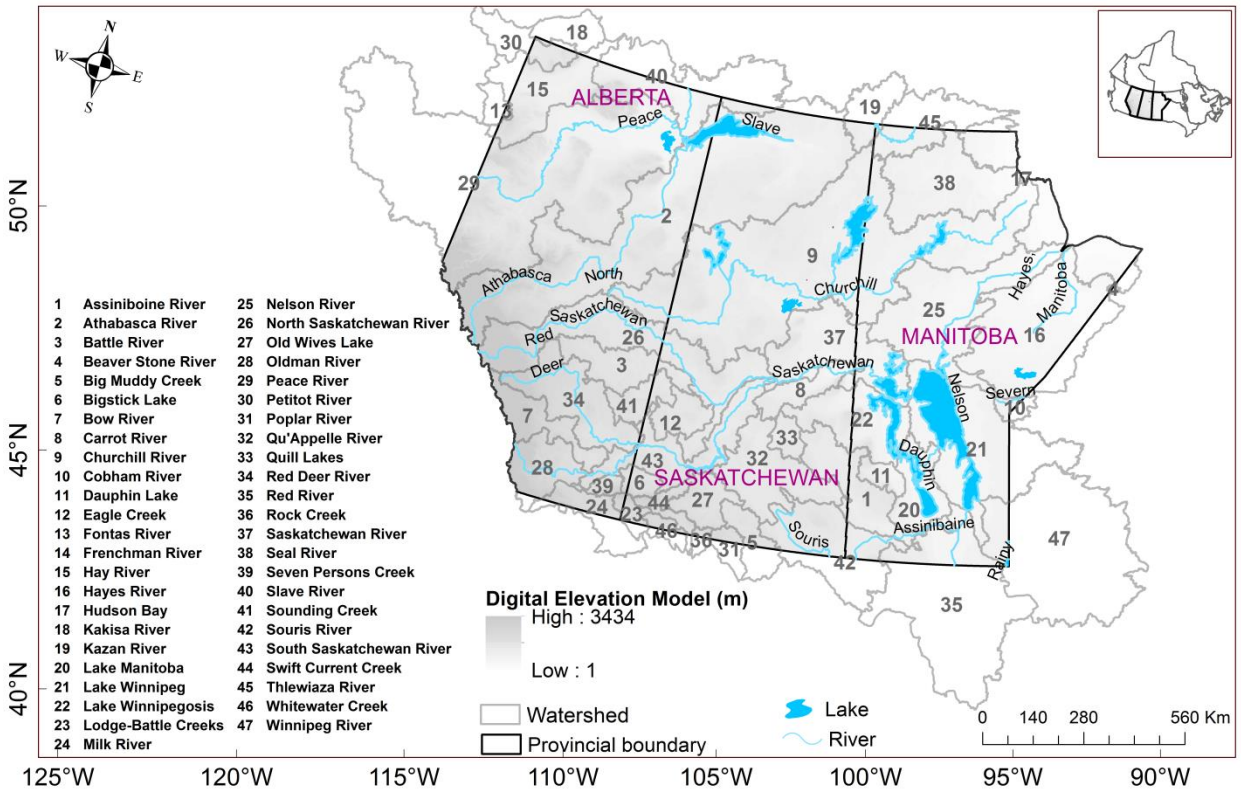

Figure 1-2: Study area showing the forty seven watersheds spanning the provinces of Alberta, Saskatchewan and Manitoba. The topography, major river systems and lakes are also indicated. The inset shows the location of the study area in Canada

1.4 Research Significance

Water resources are both important to society and ecosystems in the CPPs. To sustain health, humanity depends on a reliable and accessible clean supply of drinking water. Water is valuable not only for domestic uses, but also for its role in maintaining aquatic ecosystems and other socio-environmental sectors, including agriculture, energy production, transportation, tourism and other industrial uses. As indicated in Section 1.1, findings from the IPCC AR5

indicate that climate change due to increasing greenhouse gas (GHG) concentrations in the atmosphere continues to be a major concern.

The impacts of climate change can better be substantiated with local examples. Climate change continue to affect livelihood security in the CPPs through inducing risks and vulnerabilities in critical sectors such as health, agriculture and food security, energy, transport, water supply and sanitation, mining, industry and manufacturing, as well as other water-using sectors. For example, the recent drought of 2001–2002 was more extensive than the multi-billion dollar drought in 1988, which was felt across Canada but concentrated on the Prairies, and cost the regional economy $3.6 billion in lost agricultural production, was most likely exacerbated by a highly varying hydro-climate. Furthermore, some of the historic floods in the CPPs are directly related to high, intense rainfall events. Examples include the 2013 Alberta, 2010 Maple Creek, 2005 Cumberland House, 2000 Vanguard and 1975 Regina floods. Two additional flood events (1969 Qu'Appelle River and 1952 Eastend) are attributed to high spring runoff. Moreover, the spring and early summer of 2011 experienced severe flooding, with many Saskatchewan communities affected. At this time, winter temperatures in the Prairies were about 6.3°C above normal, leading to high spring melt and soil moisture accumulation.

Therefore, understanding these climatic changes through providing reliable future climate change projections at the scale of human social systems will allow for water resources allocators and managers to better inform their decision making processes. Particularly, knowledge of characteristics of climatic extremes such as intensity, frequency and duration is very crucial for effective and sustainable management of flood- and drought-induced risks which entail an assessment of existing infrastructure, and the design and construction of dams, reservoirs and other water supply and flood protection schemes. These local scale climatic projections will also

allow for development of new or updated operating rules for water resources availability assessments and allocation for various purposes.

This study also proposed a new framework for identifying homogeneous precipitation regions using large-scale atmospheric circulation patterns and geophysical attributes, and multivariate approaches. The study further evaluated independently the homogeneity of these regions using statistics obtained from station-based observed precipitation data at different time scales. The defined homogeneous regions were used to develop precipitation magnitude-frequency relationships of single- and multiday warm and cold season precipitation extremes through L-moments-based regional frequency analysis approach. We believe that this was the first attempt to map spatially the characteristics of seasonal precipitation extremes in the three Prairie Provinces of Canada. The proposed methodological framework constitutes a major contribution towards a better understanding of the linkages between precipitation characteristics and large scale atmospheric circulations in a regional perspective.

In any statistical downscaling investigation involving multisite multivariate predictands, it is important to preserve not only the cross-correlation structure between individual sites but also among the different predictands. However, most of the previous studies based on weather generators (WGs) tend to focus on individual sites and are therefore unable to represent the spatial structure of the observed climatic variables. However, for many water resources design and management projects, particularly in large river basins, it is important to model simultaneous sequences of multiple variables (e.g., precipitation and temperature) over large heterogeneous areas, while maintaining physically plausible spatial, temporal and inter-variable relationships. As a solution to the above issues, the GLM based WG applied in this study provided a flexible framework for accomplishing such challenging tasks. Until now, we are not aware of any

relatively simple yet effective WG that has been developed and tested to model multiple variables at 120 sites over a predominantly nonhomogeneous region.

Finally, this study was the first of its kind in Canada to perform multisite multivariate statistical downscaling over homogenous climatological regions using GLMs. Also, downscaling AOGCM outputs to daily time scale at multiple sites allows for the study of wet and dry spells, as well as climatic indices and extremes which are very critical for various adaptation and policymaking processes in different sectors of the society. The WG based approach employed in this work allowed for the generation of synthetic time series of weather data of any length which makes it suitable for use in the assessment of risks involved in the design of water resources systems.

1.5 Thesis Outline

This thesis adopts a 'dissertation by manuscript' style. Following the introductory chapter, the thesis is structured into three manuscripts, each of which is presented as a single chapter. The first manuscript (Chapter 2), "Regionalization of precipitation characteristics in the Canadian Prairie Provinces using large scale atmospheric covariates and geophysical attributes" developed a new approach for identifying homogeneous regions for regionalization of precipitation characteristics over the CPPs. This approach made use of large-scale atmospheric covariates, teleconnection indices and geographical site attributes that impact spatial patterns of precipitation to delineate homogeneous precipitation regions. The delineated regions were validated independently for homogeneity using statistics computed from observations.

The second manuscript (Chapter 3) "Multisite multivariate modeling of daily precipitation and temperature in the Canadian Prairie Provinces using Generalized Linear Models" detailed the development of a multisite multivariate stochastic modelling approach based on the GLM framework with the aim of using this framework for downscaling AOGCM outputs for climate

change impact analysis. Within this framework, temperature was modeled using a two-stage normal-heteroscedastic model while precipitation occurrence and conditional precipitation intensity processes were modelled separately. The relationship between precipitation and temperature was accounted for by using transformations of precipitation as covariates to predict temperature fields. Subsequently, large scale atmospheric covariates from the National Center for Environmental Prediction Reanalysis-I, teleconnection indices, geographical site attributes, and observed precipitation and temperature records were used to calibrate these models for the 1971–2000 period. Validation of the developed models was performed on both pre- and post-calibration period data. The developed GLMs were able to model the joint distribution of precipitation and temperature simultaneously at all sites and account for inter-variable and inter-site dependence structures.

In Chapter 4 the third manuscript ("Projected changes in precipitation and temperature characteristics over the Canadian Prairie Provinces using the Generalized Linear Model multisite multivariate statistical downscaling approach") is presented. By conditioning the developed model parameters from Chapter 3 on AOGCM outputs, future projections in terms of changes in mean and extreme states relative to the historical period were generated and analyzed. Changes in mean characteristics were investigated on both temporal and spatial scales while projected changes in climate extremes indices were studied using the extreme value theory.

Finally, in Chapter 5 a summary and the conclusions drawn from the various analyses presented in this study are provided along with some recommendations for future work.

1.6 Copyright and Author Permissions

Chapters 2 through 4 of this thesis consist of manuscripts that have been published, accepted for publication or submitted for publication. For consistency with the copyright and author rights of each publisher, proper manuscript citations are provided below. Permission to use or author rights from each publisher allowing use of the manuscripts in this thesis is included further in Appendix D.

Chapter 2: Asong ZE, Khaliq MN, Wheater HS (2015), Regionalization of precipitation characteristics in the Canadian Prairie Provinces using large-scale atmospheric covariates and geophysical attributes, *Stochastic Environmental Research and Risk Assessment*, 29 (3), 875-892.

Chapter 3: Asong ZE, Khaliq MN, Wheater HS (2015), Multisite multivariate modelling of daily precipitation and temperature in the Canadian Prairie Provinces using Generalized Linear Models. Submitted to *Journal Climate Dynamics*, paper # CLDY-D-14-00659 (status: "under review" as of 14/12/2015)

Chapter 4: Asong ZE, Khaliq MN, Wheater HS (2015), Projected changes in precipitation and temperature characteristics over the Canadian Prairie Provinces using the Generalized Linear Model multisite multivariate statistical downscaling approach. Submitted to *Journal of Hydrology, Paper # HYDROL20972* (status: "under review" as of 14/12/2015)

Contributions of the candidate: In the work presented in Chapters 2 to 4, which form the core of this thesis, the candidate (Asong, ZE) developed conceptual ideas and theoretical frameworks, carried out the analyses, and designed and prepared the manuscripts. The co-authors (Dr. Khaliq and Prof. Wheater) provided advice on various aspects of the research and critical reviews of the results and their interpretations.

References

Barnett, T. P., J. C. Adam, D. P. Lettenmaier (2005), Potential impacts of a warming climate on water availability in snow-dominated regions, *Nature Reviews*, 438 (17) doi:10.1038/nature04141.

Carter, T. R., M. L. Pary, Harasawaan, H., Nishiok, S. (1994), IPCC technical guidelines for assessing climate change impacts and adaptions, special report to Working Group II, Intergovernmental Panel on Climate Change, 59 pp.

Catell, R. B. (1966), The scree test for the number of factors. *Multivariate Behavioral Research,* 1, 245-276.

Council of Canadian Academies, (2013), Water and Agriculture in Canada: Towards Sustainable Management of Water Resources. Chapter 6 – Building the Foundation for Sustainable Management of Water in Agriculture. Available at: http://www.scienceadvice.ca/en/assessments/completed/water-agri.aspx

DeBeer, C. M., Wheater, H. S., Carey, S. K., and Chun, K. P. (2015), Recent climatic, cryospheric, and hydrological changes over the interior of western Canada: a synthesis and review, *Hydrology Earth System Science Discussions*, 12, 8615-8674, doi:10.5194/hessd-12-8615-2015

Demuth, M.N., A. Pietroniro (2003), The impact of climate change on the glaciers of the Canadian Rocky Mountain eastern slopes and implications for water resource-related adaptation in the Canadian Prairies: Phase I- Headwaters for the North Saskatchewan River Basin. Report for the Climate Change Action Fund- Prairie Adaptation Research Collaborative, 1-111.

Dingman, SL (2008), *Physical Hydrology*, second edition, Waveland Press, Inc., Long Grove, Illinois, USA.

FAO (2003), Review of world water resources by country. Water Report 23. FAO, Rome (Italy). Land and Water Development Division, 110 pp.

Fowler, H.J., Blenkinsopa, S., Tebaldi, C., (2007), Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, *International Journal of Climatology*, 27, 1547–1578.

Graham, N.E. (1990), Canonical Correlation Analysis, in WMO review of *Climate Diagnostic Methods*, World Meteorological Organization, Geneva.

Graham, S. (1988), Precipitation*: Process and Analysis*, John Wiley & Sons Ltd, Baffins Lane, Chichester, W. Sussex, UK.

Hejazi, M.I., Edmonds, J., Clarke, Kyle1, L.P., Davies, E., Chaturvedi, V., Eom, J., Wise, M., Patel, P., Calvin, K. (2013), Integrated assessment of global water scarcity over the 21st century – Part 2: climate change mitigation policies, *Hydrology and Earth System Science Discussions*, 10, 3383–3425.

IPCC (2013), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Johns, T. C., R. E. Carnell, J. F. Crossley, J. M. Gregory, J. F. B. Mitchell, C. A. Senior, S. F. B. Tett, Wood, R. A. (1997), The second Hadley Center coupled model ocean-atmosphere GCM: Model description, spin up, and validation, *Climate Dynamics*, 13, 103-134.

Kaiser, H. E. (1958), The varimax criterion for analytic rotation in factor analysis. *Pyrometrical,* 23, 187-200.

Kundzewicz, Z. W., and Coauthors (2007), Freshwater resources and their management. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. C. U. Press. Cambridge, United Kingdom: 173-210.

Murdoch, PS., JS. Baron, TL. Miller (2000), Potential effects of climate change on surface-water quality in North America, *American Water Resources Association,* 36 (2), 347–366.

Natural Resources Canada (2010), http://www.nrcan.gc.ca/earth-sciences/climate-change/community-adaptation/642[accessed June, 2015].

Prairie Adaptation Research Collaborative (2010), (http://www.parc.ca/saskadapt/extreme-events/thunder-flood [accessed June, 2015].

Preisendorfer, R. W., and Mobley, C. D. (1988), *Principal Component Analysis in Meteorology and Oceanography.* Amsterdam, Elsevier, 425 pp.

Richman, M. B. (1986), Rotation of principal components (review article), *International Journal of Climatology,* 6, 293-335.

Sauchyn, D., Kulshreshtha, S. (2008), Climate change impacts on Canada's Prairie Provinces: A summary of our state of knowledge, from "Prairies" in From Impacts to

Adaptation: Canada in a changing Climate 2007, edited by D. Lemmen et al., Government of Canada, Ottawa.

Stewart, R., RG. Lawford, A. Boisvert (2011), The 1999-2005 Canadian prairie drought: science, impacts, and lessons. Winnipeg, Manitoba, Drought Research Initiative

Stocker, T.F., and Coauthors (2013), Technical Summary. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Viessman, W., Lewis, G.L. (2011), Introduction to Hydrology, fifth edition, Pearson Education, Inc., Upper Saddle River, New Jersey, USA.

Vincent, L. A., X. Zhang, R. D. Brown, Y. Feng, E. Mekis, E. J. Milewska, H. Wan, X. L. Wang (2015), Observed Trends in Canada's Climate and Influence of Low-Frequency Variability Modes, *Journal of Climate*, 28, 4545–4560.

von Storch, H., Burger, G., Schnur, R., and von Storch, J. S. (1995), Principal oscillation pattern: A review, *Journal of Climate*, 8, 377–400.

Vörösmarty, CJ, P. Green, J. Salisbury, RB. Lammers (2000), Global Water Resources: Vulnerability from Climate Change and Population Growth, *Science*, 289 (5477), 284-288.

Wilks, D.S. (2011), *Statistical Methods in the Atmospheric Sciences*, 3rd Edition, Academic Press, London.

Wilks, S.D. (1995), *Statistical Methods in the Atmospheric Sciences*: International Geophysics Series, Volume 59. Academic Press, San Diego, CA.

Young, RA. (2005), Determining the Economic Value of Water: Concepts and Methods. Resources for the Future, Washington, DC, USA.

Zhang, X., LA. Vincent, W.D. Hogg, A. Niitsoo (2000): Temperature and precipitation trends in Canada during the 20th century, *Atmosphere-Ocean*, 38 (3), 395-429, doi: 10.1080/07055900.2000.9649654.

CHAPTER 2
REGIONALIZATION OF PRECIPITATION CHARACTERISTICS IN THE CANADIAN PRAIRIE PROVINCES USING LARGE SCALE ATMOSPHERIC COVARIATES AND GEOPHYSICAL ATTRIBUTES

As discussed in Chapter 1, global climate change is expected to continue to have serious implications on the earth's ecosystems, particularly its impacts on water resources and the human environments. To cope with this situation, proper mitigation and adaption strategies geared towards minimizing risks and vulnerabilities in water-sensitive sectors such as aquatic ecosystems and other socio-environmental sectors, including agriculture, energy production, transportation, tourism and the ever-changing role of the urban water sector must be developed. For example, urban environments are characterized by human encroachments on natural drainage systems, impervious pavements, and immense settlement. Hence, these environments are vulnerable to flash flooding resulting mostly from extreme rainfalls, exacerbated occasionally due to rain on snow events. Therefore, improved estimation of design storms by incorporating the effect of climate change is indispensable for use in the design, operation and maintenance of urban water infrastructure such as pipes, storm sewers, retention and detention ponds, and culverts.

To this end, the problem often faced by practitioners is that in some regions of interest, there may be no measurement stations available or the observed record length is too short. As a potential remedy, regional/pooled frequency analysis, whereby information from sites within the entire region is utilized to improve at-site estimates, has been widely adopted. The pooling process is usually based on dividing a larger region into smaller sub-regions with homogeneous characteristics of the variable of interest. In the literature, the established approaches tend to rely heavily on statistics computed from observed data rather than the large scale climatic variables that influence regional and local weather patterns at various temporal and spatial scales.

In this Chapter, a new approach for identifying homogeneous regions for regionalization of precipitation characteristics was proposed for the Canadian Prairie Provinces. This approach made use of large-scale atmospheric covariates, teleconnection indices and geographical site attributes that impact spatial patterns of precipitation to delineate homogeneous precipitation regions. The delineated regions were validated independently for homogeneity using statistics computed from observations. These homogeneous regions served the basis for developing a multisite multivariate stochastic modelling approach for precipitation and temperature fields, presented in Chapter 3. This chapter contains the following published journal paper:

1. **Asong ZE**, Khaliq MN, Wheater HS (2015), Regionalization of precipitation characteristics in the Canadian Prairie Provinces using large-scale atmospheric covariates and geophysical attributes, *Stochastic Environmental Research and Risk Assessment*, 29 (3), 875-892.

Abstract

Observed data at most stations are often inadequate to obtain reliable estimates of many hydro-meteorological variables that not only define water availability across a region but also the vulnerability of social infrastructure to climatic extremes. To overcome this, data from neighboring sites with similar statistical characteristics are often pooled. The pooling process is based on partitioning of a larger region into smaller sub-regions with homogeneous features of interest. The established approaches rely heavily on statistics computed from observed precipitation data rather than the covariates that play a significant role in modulating the regional and local climate patterns at various temporal and spatial scales. In this study, a new approach for identifying homogeneous regions for regionalization of precipitation characteristics was proposed for the Canadian Prairie Provinces. This approach incorporated information about large-scale atmospheric covariates, teleconnection indices and geographical site attributes that

impact spatial patterns of precipitation in order to delineate homogeneous precipitation regions through combined use of multivariate approaches – Principal Component Analysis, Canonical Correlation Analysis and Fuzzy C-means clustering. Results of the analyses suggested that the study area can be partitioned into five homogeneous regions. These partitions were validated independently for homogeneity using statistics computed from monthly and seasonal precipitation totals, and seasonal extremes from a network of observation stations. Furthermore, based on the identified regions, precipitation magnitude-frequency relationships of warm and cold season single- and multi-day precipitation extremes, developed through regional frequency analysis, were mapped spatially. Such estimates are important for numerous water resources related activities.

Keywords: Large-scale covariates; Canadian Prairie Provinces; multivariate approaches; homogeneous regions; seasonal precipitation extremes

2.1 Introduction

The role of large-scale atmospheric circulation and teleconnection patterns in modulating regional hydro-climate at various temporal and spatial scales has been examined long ago. For example, Burger (1958) studied the relationships between atmospheric circulation patterns and local-scale weather observations (including mean, maximum and minimum daily temperatures, precipitation amounts and cloudiness) from 1890 to 1950 at four German cities (Berlin, Bremen, Karlsruhe and Munich). He found good agreement between surface climatic variables and general atmospheric circulations. Also, Lamb (1977) found that local scale precipitation is strongly linked to atmospheric circulation patterns.

Loikith and Broccoli (2012) examined characteristics of observed atmospheric circulation patterns associated with temperature extremes over North America. They found warm extremes at most locations to be associated with positive 500-hPa geopotential height and sea level

pressure anomalies. In another study, Ropelewski and Halpert (1986) found that above normal precipitation and positive temperature anomalies in North America are associated with El Niño/Southern Oscillation (ENSO). The influence of atmospheric and oceanic variability associated with growing season droughts and pluvials in the Canadian Prairies, western Canada, has been studied by Shabbar et al. (2011). Findings from the analyses indicate that moisture from the Gulf of Mexico is notably decreased during the identified drought seasons. Stronger than normal subsidence associated with anomalously high pressure over north-western North America also leads to weakened moisture transport from the Pacific Ocean. Conversely, during pluvial seasons, low-level flow aided by the circulation associated with increased cyclone frequency over western North America brings abundant moisture northward into the southern Prairie region.

Various modes of low-frequency circulation variability such as the Pacific Decadal Oscillation (PDO), Pacific North American (PNA) pattern, Artic Oscillation (AO), North Pacific (NP) pattern and the North Atlantic Oscillation (NAO) have been established as influencing North American hydro-climate and streamflow characteristics. For example, Terray and Cassou (2000), Mantua and Hare (2002), Sheridan (2003), Romolo et al. (2006a), Woo and Thorn (2008), Ge and Gong (2009), Ghatak et al. (2010), Khaliq and Gachon (2010), Perez-Valdivia et al. (2012) and Zhao et al. (2013) found significant correlations between teleconnection indices and hydro-climatic variability for northwestern North America.

Regionalization of precipitation characteristics has many applications in a number of fields of socio-economic importance such as planning of agriculture, watershed management, and investigating and understanding temporal and spatial structures of extreme events. For regionalization, reliable estimates of various precipitation characteristics require availability of long records of historical measurements at numerous stations within a region. However,

observed precipitation data at local scales are often inadequate to facilitate such estimates. To overcome this problem, observations from neighboring sites with like statistical characteristics are pooled (e.g., Stedinger et al. 1993; Rossi and Villani, 1994; Bobée and Rasmussen, 1995; Robson and Reed, 1999). In this study, regionalization refers to the procedure involved in identifying homogeneous sub-regions of a larger region consisting of sites having similar statistical as well as climatological characteristics.

Several studies have been carried out worldwide to identify homogeneous regions consisting of sites with similar hydrological and physical characteristics (e.g. Mosley 1981; Gottschalk 1985; Burn 1989). The classical approaches include cluster analysis (CA) (Alila 1999), L-moments based approaches (e.g., Guttman et al. 1993; Hosking and Wallis, 1997), principal component analysis (PCA) (Ehrendorfer, 1987; Basalirwa, 1995; Comrie and Glenn, 1998), and PCA in association with CA (e.g., Baeriswyl and Rebetez, 1997; Wu et al., 2006). Other regionalization approaches include the region of influence approach (Burn 1988, 1990a, 1990b), entropy theory (Rianna et al., 2012), K-means clustering (Satyanarayana and Srinivas, 2008; Cosmo et al., 2011; Kannan and Ghosh, 2011), fuzzy cluster analysis (Rao and Srinivas, 2006b; Satyanarayana and Srinivas, 2011), self-organizing maps (Lin and Chen, 2006), elementary linkage analysis (McQuitty 1957), canonical correlation analysis (CCA) (Cavadias, 1989, 1990; Ouarda et al., 2001) and spatial correlation analysis (Gadgil et al., 1993).

Apart from the recent studies of Satyanarayana and Srinivas (2008, 2011), who utilized large-scale atmospheric variables from the National Centers for Environmental Prediction/National Centre for Atmospheric Research (NCEP/NCAR) reanalysis products to regionalize precipitation climates in India, the importance of large-scale atmospheric circulations and their impact on regional precipitation patterns has had limited exploitation for regionalization of precipitation climates. Also, in the UK, Fowler et al. (2000) classified the

region of Yorkshire into distinct precipitation regions, using the objective Lamb weather types (Lamb, 1972) and daily precipitation statistics.

In Canada, with respect to regionalization, Adamowski et al. (1996) performed an analysis of regional rainfall patterns based on homogeneous rainfall regions delineated by means of L-moments. Following this, Alila (1999) used a hierarchical CA approach in conjunction with L-moments for regionalization of annual maxima of precipitation. Results of Monte Carlo simulations indicated that design storms estimated by the proposed hierarchical approach are substantially more accurate than those estimated by single-site analyses.

However, such approaches are circumscribed because they rely heavily on statistics, computed from observed precipitation records. Also, long records of precipitation data and high station density are often required for reliable delineation of homogeneous regions using these approaches. In addition, the statistics used for delimiting homogeneous regions are also used for testing purposes (e.g., Burn, 1988; Adamowski et al., 1996; Alila, 1999). On the basis of these observed data-driven approaches, independent validation of delineated regions is not generally possible.

The methodology proposed in this study differs from others in that it utilized not only atmospheric covariates and geographical site attributes but also considered indices of teleconnection patterns. The suggested framework also facilitated an independent validation of identified homogeneous regions at different time scales using statistics computed from observed precipitation records. To the authors' knowledge, until now, such an inclusive regionalization approach did not exist. In summary, the goal of this research was to explore and document the utility of large-scale atmospheric circulation patterns in delimiting homogeneous precipitation climates in the Canadian Prairie Provinces of Alberta, Saskatchewan and Manitoba through multivariate approaches (PCA, CCA and Fuzzy C-Means (FCM) clustering) and to evaluate

independently the homogeneity of these regions using statistics obtained from station-based observed precipitation data at different time scales. The homogeneous regions defined so were used to develop precipitation magnitude-frequency relationships of single- and multiday warm (JJA) and cold (DJF) season precipitation extremes through L-moments-based regional frequency analysis (RFA) approach. We believe that this is the first attempt to map spatially the characteristics of seasonal precipitation extremes in the three Prairie Provinces of Canada, which serves as an additional motivation for the present study.

The remainder of this paper is structured as follows: Section 2.2 describes the datasets used. The proposed methodology for identifying homogeneous precipitation climates is given in Section 2.3, alongside description of a framework for developing precipitation magnitude-frequency relationships for seasonal precipitation extremes. Results of the study are presented and discussed in Section 2.4, while a summary and conclusions are given in Section 2.5.

2.2 Datasets Used

The datasets used in this study include annual, monthly, and seasonal (winter and summer) precipitation totals and seasonal precipitation extremes, derived from daily precipitation observations, for the 1961 to 2005 period from a network of 120 stations (Figure 2-1), obtained from Environment Canada (http://www.ec.gc.ca). In winter, precipitation extremes are dominated by snow, while in summer, extremes are dominated by rain. Considering both cold and warm season precipitation extremes will provide an opportunity to validate the delimited regions across two different seasons (winter and summer). The underlying daily precipitation dataset is categorized as the *Second Generation of Daily Adjusted Precipitation Data for Canada*. The impact of the adjustments on daily, monthly and annual precipitation totals was examined in detail by Mekis and Vincent (2011). On an annual time scale, the south-central and northwestern parts of the study area generally experience the least amount of precipitation. On a

seasonal basis, in summer (JJA), the south-central and northern parts tend to experience drier conditions relative to other parts. While in winter (DJF), the south-central to southwestern and southern parts of Manitoba tend to be drier compared to the rest of the study area.

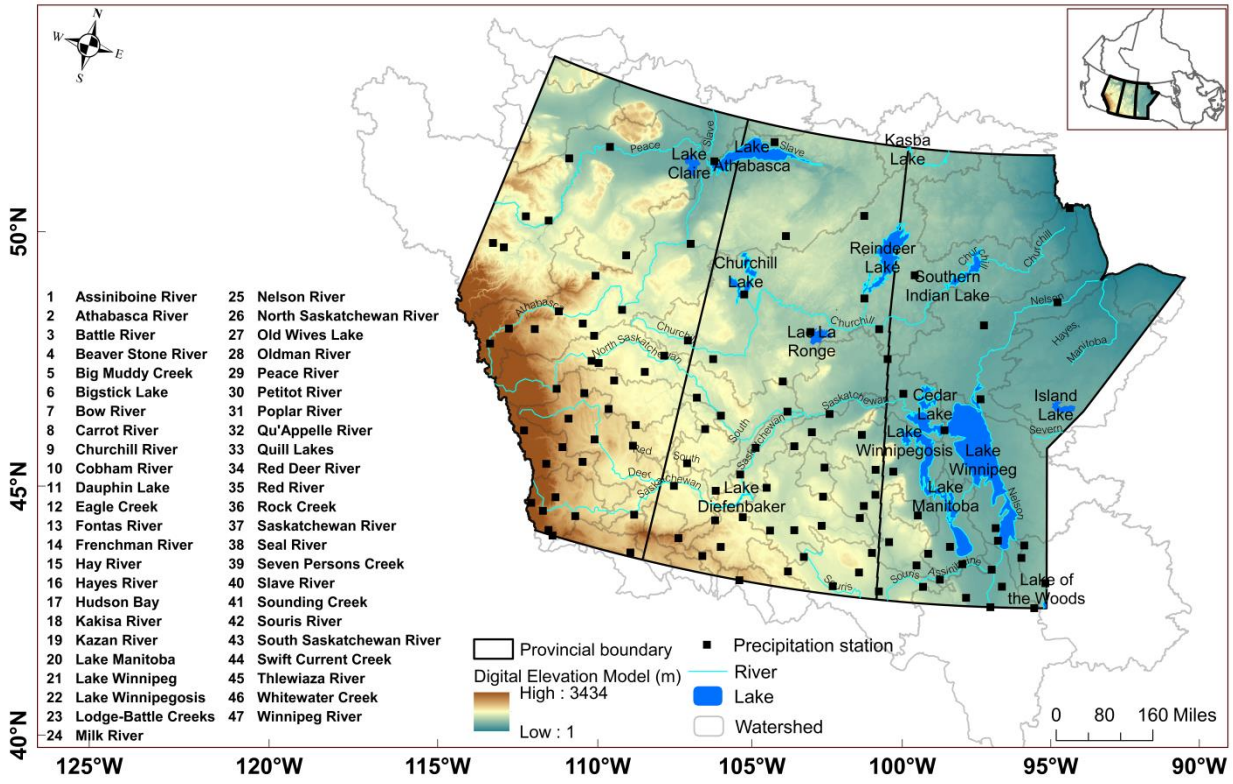

Figure 2-1: Map of the study area (provinces from left to right – Alberta, Saskatchewan and Manitoba) showing different watersheds and the location of precipitation stations used in the study. List of 47 watersheds covering the study area is also provided. Inset shows location of the study area in Canada

Daily values of large-scale atmospheric covariates were derived from the NCEP reanalysis-I dataset (Kalnay et al., 1996) for the 1961–2005 period over a spatial domain of 65 grid squares encompassing latitude 47^{o}N–61^{o}N and longitude 121^{o}W to 88^{o}W. In total, 21 large scale covariates (wind speed at 10-m, 500- and 850-hPa; U-component and V-component at 10-m, 500- and 850-hPa, vertical velocity, geo-potential height, specific humidity, and relative humidity at 850- and 500-hPa; total cloud cover, mean sea level pressure, precipitable water and 2-m air temperature) were explored. Monthly values of indices of teleconnection patterns were

sourced from the Joint Institute for the Study of the Atmosphere and Ocean, University of Washington (http://jisao.washington.edu/analyses0302/) and Climate Research Unit, University of East Anglia (http://www.cru.uea.ac.uk/cru/data/soi/). Though the current framework only employed atmospheric fields within the domain of the study area as inputs, it is worthy to note that the large-scale atmospheric features which influence precipitation within the study area could be located well outside the domain of interest (e.g., Szeto et al., 2011). However, such possibilities are not explored in this study. The above mentioned datasets were re-expressed by extracting the long-term mean for the given station/grid and dividing by the respective standard deviation (von Storch, 1995 and Wilks, 2011). The resulting dimensionless standardized anomalies were used as inputs for PCA and CCA.

2.3 Methodology

Multivariate approaches such as PCA and CCA were applied to study precipitation characteristics and to understand the relationships between precipitation amounts and large-scale atmospheric covariates, indices of teleconnection patterns and geographical site attributes. The PCA and CCA (see sub-section 2.3.1) were also utilized to screen covariates and indices of teleconnection patterns to form feature vectors which served as input to the FCM clustering algorithm for defining soft clusters (see sub-section 2.3.2). Next, the regions identified were independently validated and re-adjusted to improve their statistical homogeneity and physical consistency through L-moments-based heterogeneity tests using monthly and seasonal precipitation totals, and single- and multiday seasonal precipitation extremes (see subsection 2.3.3 and 2.3.4). Finally, an approach based on the RFA algorithm (Hosking and Wallis 1997) was described for developing precipitation magnitude-frequency relationships of 1-, 3- and 5-day seasonal precipitation extremes (see sub-section 2.3.5).

2.3.1. Predictor Selection using PCA and CCA

Several approaches to predictor selection abound in the literature among which include multiple linear regression (Maheras et al., 2004), the Sampson correlation ratio to predictors (Tatli et al., 2004), empirical orthogonal functions (EOFs) and CCA (Xoplaki et al., 2004).

Given monthly precipitation for 45 years at 120 stations in the study area, we have $n = 540$ observations for $p = 120$ stations. In many cases, the observations at p stations are highly correlated, particularly if the stations are in close proximity. It is reasonable to select a few variables $m < p$, which express most of the information contained in the original $n * p$ matrix, **Z**, and are also uncorrelated with each other (Richman 1986; Graham 1988; Preisendorfer and Mobley, 1988; Wilks 2011).

In order to isolate the major modes of inter-annual variability in the monthly precipitation totals, the normalized time-space dataset is subject to S-mode PCA to retain *m* Principal Components (PCs) or EOFs, which explain maximum variation in the original data. Subsequently, indices of teleconnection patterns were correlated with the leading *m* PCs of precipitation and indices with statistically significant (at $\alpha = 5\%$ significance level) relationships were retained and used as plausible precipitation-sensitive attributes/feature vectors for the FCM clustering algorithm.

Additional feature vectors were formed through investigating spatial and temporal inter-relationships between 21 NCEP candidate predictors (*X*-field) and monthly precipitation (*Y*-field) by subjecting the two fields to CCA. As *X* and *Y* contain many stations/grids, the data were reconstructed to reduce their dimension, as well as redundancy using the PCA. This procedure was applied to *X* and *Y* separately and the leading *m* significant PCs which accounted for > 90% of the variance in each case were retained. CCA was then applied to the leading *m* PCs of *X* and *Y* in order to isolate the uncorrelated dominant pairs/patterns of co-variability between them.

Geographic site attributes such as latitude, longitude and elevation were correlated with precipitation using a similar CCA model in order to isolate and retain only the dominant modes of co-variability between the two fields, thereby reducing the issue of multi-collinearity.

The significance of the canonical correlation between *X* and *Y* was tested at 95% confidence level using, Rao's approximate *F* statistic and Bartlett's *chi-square test* (Krzanowski, 1988). Using this approach, only statistically significant and physically meaningful CCA patterns were retained as additional attributes for the FCM clustering algorithm.

2.3.2. Delineation of Soft Clusters using the FCM Clustering Algorithm

Hydro-meteorological characteristics of some parts of the study area are partly similar, thus it is difficult to justify assigning a site solely to one sub-region with a hard boundary. Fuzzy clustering allows a site to have partial or distributed membership in more than one sub-region. Thus, only soft clusters with vague boundaries are formed unlike hard clusters that one would obtain using the K-means method (McQueen, 1967), though this condition was relaxed for developing single- and multiday precipitation magnitude-frequency relationships. The FCM algorithm used in this study is based on the work of Dunn (1974) and Bezdek (1981). This algorithm uses iterative optimization of a fuzzy objective function in order to subdivide *N* sites in a region into *c* fuzzy soft clusters.

For *N* sites in the study region, *n* attributes, which impact precipitation at each site, are identified. Let $y_i = \left[y_{1i}, ..., y_{ji}, ..., y_{ni} \right]^{'} \in \Re^n$, denote the *i*th feature vector depicting the *i*th site in *n*-dimensional attribute space. Here, y_{ji} is the *j*th attribute in the feature vector y_i. For those attributes which strongly modulate precipitation patterns in an area and have greater variance than the least important ones, the feature vectors must be rescaled using Equation 2-1.

$$x_{ji} = \frac{\left(y_{ji} - \bar{y}_j\right)}{\sigma_j} \quad \text{for } 1 \le i \le N; \quad 1 \le j \le n \tag{2-1}$$

where x_{ji} denotes the rescaled value of y_{ji}, σ_j represents the standard deviation of the attribute *j* and $\bar{y}_j$ is the mean value of the attribute *j* over all *N* feature vectors. Consider an $n \times N$ matrix, **Q**, containing the rescaled feature vectors:

$$\mathrm{Q} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,N} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,N} \end{bmatrix} \tag{2-2}$$

The FCM algorithm partitions **Q** into *c* overlapping fuzzy clusters by minimizing the objective function and constraints described in Bezdek (1981). By varying c and μ (fuzzifier or weight exponent), different preliminary sets of clusters are obtained. Subsequently, an optimal number of clusters, using the Xie-Beni fuzzy cluster validity index (Xie and Beni, 1991), is determined. Finally, to form fuzzy clusters, a site was assigned to the cluster(s) in which it has membership more than or equal to a threshold T_i, computed following guidelines from Srinivas et al. (2008) and Satyanarayana and Srinivas (2011):

$$T_i = \max\left\{\frac{1}{c}, \frac{1}{2}\left[\max_{1 \le k \le c}\left(\mu_{ik}\right)\right]\right\} \tag{2-3}$$

The fuzzy clusters, identified using the above mentioned approach, need to be evaluated for statistical homogeneity. For that, the homogeneity tests and discordancy measures proposed by Hosking and Wallis (1997) were applied to observed precipitation sequences and not to feature vectors from which the clusters are identified. The detailed procedure is described below.

2.3.3. Screening of Data and Adjustment of Clusters using the Discordancy Measure

The discordancy measure *D* was used for detecting inconsistencies and outliers in the observed data and for adjusting (if necessary) soft clusters, formed using the regionalization

attributes (i.e. atmospheric variables, teleconnection indices and geographical site attributes). Observed precipitation data from 120 stations/sites were screened. A site was judged as discordant if *D* equals or exceeds the critical value, which is generally taken to be 3 (see Hosking and Wallis, 1997). A soft cluster was re-adjusted by eliminating those sites that are grossly discordant with respect to the majority of the sites in the same cluster. A successfully delineated fuzzy soft cluster/region is denoted by *R* and the one re-adjusted by R^{adj}. Consequently, if a site *i* belongs to $R_{(i)}^{adj}$, then its membership in each of the $R_{(i)}^{adj}$ regions is updated using Equation 2-4:

$$\mu_{ik}^{adj} = \frac{\mu_{ik}}{\sum_{k=1}^{R_{(i)}^{adj}} \mu_{ik}} \text{ for } 1 \le k \le c, \quad 1 \le i \le N \tag{2-4}$$

where $\mu_{ik}^{adj} \in [0,1]$ represents the updated membership of site *i* in the *k*th fuzzy region. Finally, if a site completely belongs to one fuzzy region, the membership of that site in that fuzzy region has to be updated to 1, and to 0 in all other regions.

2.3.4. Validation of Soft Clusters using the Heterogeneity Test

The re-adjusted fuzzy regions identified using the foregoing methodology were validated independently for statistical homogeneity using statistics computed from monthly and seasonal precipitation totals, and seasonal precipitation extremes. The homogeneity test is centered on the idea that if a region is homogeneous, then all sites should have the same population L-moment ratios (LMRs), though their sample LMRs might differ owing to sampling variability. Using the observed dataset distributed among fuzzy regions defined according to the procedure described in sub-section 2.3.2, regional average LMRs, i.e. L-coefficient of variation (L-CV), L-skewness and L-kurtosis, were computed. Then, the $H_k, k = 1,2,3,$ statistics that include the weighted standard deviation of at-site sample L-CV (H_1), the weighted average distance from the site *i* to

the group weighted mean in the 2-dimensional space of L-CV and L-skewness (i.e., H_2), and the weighted average distance from the site *i* to the group weighted mean in the 2-dimensional space of L-skewness and L-kurtosis (i.e., H_3) were calculated. Subsequently, a kappa distribution is fit to the regional average LMRs. To determine what would be expected by chance, 1,000 Monte Carlo simulations of a homogeneous region with sites having record lengths equal to those of the observed data were performed. For each of the simulated 'homogeneous' regions obtained using the kappa distribution, V_k statistics were calculated and the mean (μ_{vk}) and standard deviation (σ_{vk}) of these statistics were obtained. Finally, the heterogeneity measures H_k were determined. A region is considered as "acceptably homogeneous" if $H_k < 1$, "possibly heterogeneous" if $1 \le H_k < 2$ and "definitely heterogeneous" if $H_k \ge 2$ (Hosking and Wallis, 1997).

2.3.5. Regional Frequency Analysis (RFA) of Single- and Multiday Precipitation Extremes

2.3.5.1. Tests for stationarity and autocorrelation

After defining homogeneous precipitation climates following the approach described above, RFA of 1-, 3- and 5-day cold (DJF) and warm (JJA) season precipitation extremes was carried out to develop precipitation magnitude-frequency relationships corresponding to selected return periods (i.e. 5-, 20- and 100-year). Prior to applying the RFA algorithm, it was important to ensure that the basic governing assumptions of stationarity and serial independence were satisfied for performing RFA. To test stationarity, firstly, the rank-based Mann-Kendall trend test (Mann, 1945; Kendall, 1975) was applied. Next, the modified Mann-Kendall test (Hamed and Rao, 1998) which accounts for serial correlation was also applied. Statistical significance of trends was evaluated at the 5% significance level. The serial independence assumption was

checked by examining lag-1 autocorrelation coefficients of the samples of seasonal extremes at the 5% significance level, following the guidelines from Wilks (2011).

2.3.5.2. Choice of a regional frequency distribution and estimation of regional growth curves

The RFA algorithm was applied to defuzzified homogeneous regions, unlike the soft clusters identified following the approach described in sub-section 2.3.2. The defuzzification was carried out to select a single regional frequency distribution for each region separately and to avoid the issues related to the presence of sites with fractional memberships in various regions. To achieve this, the fuzzy partition matrix was defuzzified or hardened using the maximum membership method proposed in Hall and Minns (1999).

For selecting a regional frequency distribution from a number of probable candidates, various goodness-of-fit techniques have been proposed in the literature. In this study, we used the LMRs diagram, which compares the relationship of L-kurtosis versus L-skewness for various commonly used distributions with the corresponding relationship obtained from the at-site and regionally averaged L-moments, in conjunction with the Z-statistic. The *Z*-statistic measures how well the theoretical L-kurtosis of a fitted distribution matches the regional average L-kurtosis of the observed data (Hosking and Wallis, 1997). The fit of the distribution is considered satisfactory if $|Z| \leq 1.64$, which roughly corresponds to an acceptance of the hypothesized distribution at a confidence level of 90%. The list of distributions considered include GLO: Generalized Logistic, GNO: Generalized Normal, GPA: Generalized Pareto, PE3: Pearson Type III, GEV: Generalized Extreme Value, and Wak: Wakeby.

A regional growth curve represents a dimensionless relationship between frequency (or return period) and magnitude of extreme values. Estimation of the regional growth curves is generally carried out using the index-flood procedure, which is based on the concept that mean

(or median) standardized at-site data within a homogeneous region follows the same frequency distribution. Here, at-site mean is taken as the standardizing constant and the regional growth curves are estimated following the L-moments-based method of Hosking and Wallis (1997). For the purpose of visualization, various return levels of seasonal precipitation extremes are spatially interpolated using the ordinary kriging method (Oliver, 1990) to portray the spatial structure of selected return values of seasonal precipitation extremes over the study area. The regionalization algorithm developed in this study was implemented in R and MATLAB, where both the PCA and CCA algorithms are readily available. For the FCM algorithm, Fuzzy Logic Toolbox from MATLAB was used.

2.4 Results and Discussion

In this section, detailed results of the study are presented and discussed. First, the results of the attributes selection procedure are presented in sub-section 2.4.1. Following this, the FCM clustering-based identified fuzzy soft regions are described and discussed in sub-section 2.4.2. Next, results of an independent validation of the fuzzy regions are presented in sub-section 2.4.3. Finally, the results of RFA algorithm applied to single- and multiday seasonal precipitation extremes are demonstrated in sub-section 2.4.4.

2.4.1 Precipitation-Sensitive Large-Scale Atmospheric Covariates, Teleconnection Patterns and Geographical Site Attributes

Prior to the selection of potential candidate predictors, PCA was applied to monthly precipitation series from a network of 120 stations distributed across the study area. This initially led to 10 leading and physically meaningful PCs, which together accounted for 64.89% of the original variability in precipitation, being retained for further analysis. However, to realize a more parsimonious model, eight PCs, which together accounted for 58.75% of the variance, were used to select possible covariates. Figure 2-2 shows the spatial patterns of the 8 leading PCs

along with their explained variances. Concerning the component scores, the first PC (with 22.76% explained variance) has entirely positive scores. It shows positive scores with strong signals mainly in the southern-central part of the study area and weak positive signals elsewhere. Stronger positive scores mean a tendency toward higher variability in precipitation amount and weaker scores mean lower variability. The second PC (with 9.56% explained variance) captures a component of variability mainly in the northwest where a strong positive signal predominates, while the third PC explains 6.28% of the variability in the southwest, close to the foothills of the Rocky Mountains. The fourth PC has a strong positive signal to the southeast and accounts for 5.03% of the original variability of monthly precipitation. Obviously, subsequent PCs explained a lesser amount of variability. The fifth to eighth PCs each explained less than 3% of the variance and give loadings with several small areas of coherent variability. Note that none of the individual PCs explained a significant amount of variability in precipitation, with over a third of the variance unaccounted for altogether. It is possible that monthly precipitation patterns in the study area are much more heterogeneous, involving coupling or interaction between indices of teleconnection patterns, large-scale atmospheric covariates and site attributes.

In order to determine possible large-scale forcing of the observed variability in precipitation, the retained eight PCs of monthly precipitation were correlated with monthly indices of teleconnection patterns (Table 2-1) and only those indices which are significantly correlated with any of the PCs were considered to form feature vectors (i.e. inputs) for the FCM clustering algorithm. Results shown in Table 2-1 indicate that the PNA index is significantly correlated with PC3 (i.e. positive scores to the southwest of the study area; Figure 2-2) and PC5 (positive scores to the northeast of the study area; Figure 2-2). The PDO index is also significantly correlated with PC5. Statistically significant relationships are found between PC5 and PDO values up to 2 prior months. Similarly, PC3 is significantly related with the prior

month's PNA index. Other indices were not found significantly correlated with any of the retained PCs. Thus, the indices of PNA and PDO constituted the first two feature vectors for the FCM clustering algorithm.

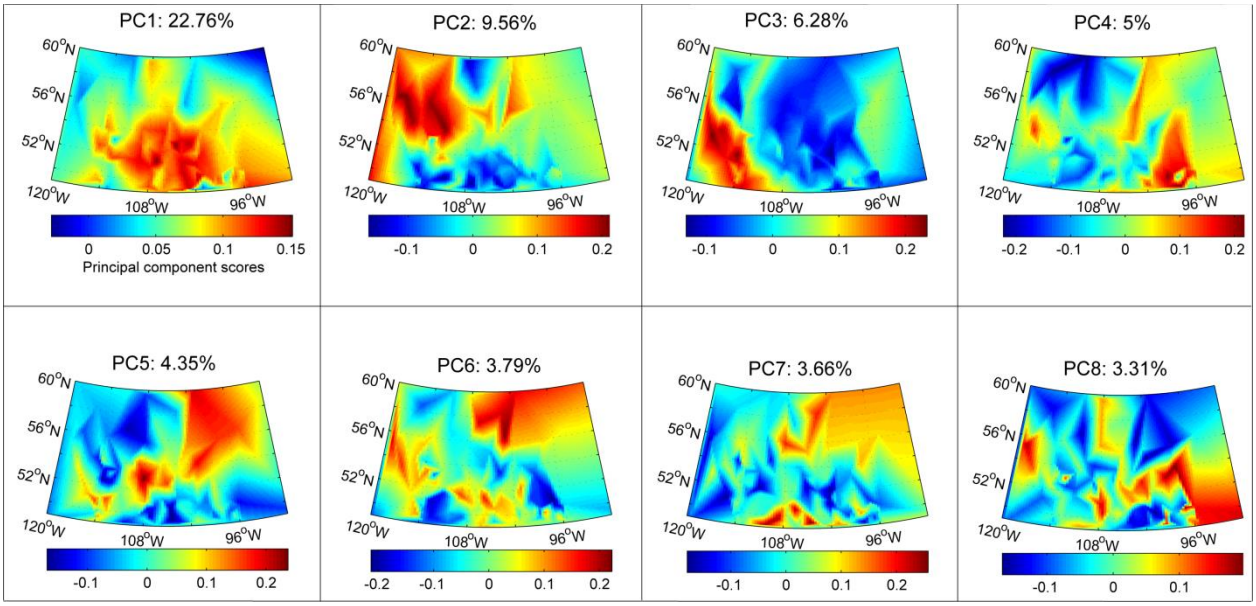

Figure 2-2: Spatial patterns of first eight PCs of monthly precipitation over the study area. Percent explained variance is indicated at the top of each panel, while the component scores are shown at the bottom of each panel

Table 2-1: Relationship between indices of teleconnection patterns and eight leading PCs of monthly precipitation. Significant (at $\alpha = 5\%$ significance level) relationship is shown in bold. Values within parenthesis represent attained significance in the case of indices and explained variance in the case of PCs. Indices of AO, NAO and NP are not found significantly correlated with any of the PCs of monthly precipitation

PCs	PNA	AO	NAO	NP	PDO
PC1 (22.7)	-0.03 (0.84)	*	*	*	-0.07 (0.63)
PC2 (9.60)	-0.16 (0.27)	*	*	*	-0.07 (0.66)
PC3 (6.28)	**-0.43 (0.03)**	*	*	*	-0.05 (0.74)
PC4 (5.03)	0.05 (0.71)	*	*	*	-0.19 (0.20)
PC5 (4.36)	**-0.36 (0.04)**	*	*	*	**-0.35 (0.01)**
PC6 (3.80)	0.07 (0.62)	*	*	*	0.09 (0.50)
PC7 (3.67)	0.26 (0.07)	*	*	*	0.16 (0.28)
PC8 (3.31)	-0.06 (0.66)	*	*	*	-0.06 (0.67)

These findings are consistent with previous studies. For example, Bonsal and Lawford (1999) found significant relations between El Niño and La Niña events and summer extended dry spells

over the Canadian Prairies that lie in southern parts of the study area. Brown and Goodison (1996) found that reduced snow cover over western Canada was associated with the positive phase of the PNA teleconnection pattern. Romolo et al. (2006a) related snow accumulation and melt in the Peace River basin with the Southern Oscillation Index and PNA.

The relationships between the 21 NCEP candidate predictors and precipitation were investigated by subjecting them to CCA. From this analysis, eight surface to upper level significant covariates namely 2-m air temperature, 850-hPa relative humidity, 500-hPa specific humidity, precipitable water, mean sea level pressure, horizontal wind components (850-hPa meridional and 10-m zonal wind), and vertical velocity (i.e., omega at 500-hPa) were found as probable candidate precipitation-sensitive climate attributes.

Figure 2-3 shows the spatial pattern of first CCA pairs (i.e., CCA1) for each predictor variable with statistically significant heterogeneous canonical correlation with the corresponding spatial pattern in the precipitation field. For example, the first canonical spatial pattern for relative humidity showed positive correlations across the northwestern part and explained 41.3% of the variance, while the corresponding canonical pattern explained 26.2% of the variance in the precipitation field. This relationship resulted in a dipole-like structure with positive and negative correlations mostly to the northern and southern parts, respectively. This CCA pair for the precipitation field has characteristics which are similar to PC2 of precipitation shown in Figure 2-2. It can then be deduced that the precipitation variability in the northwest and parts of the northeast of the study area is probably linked to a component of variability in the relative humidity at the monthly timescale. It is possible that two minor components might correlate very highly, while the explained variance of the variables is very low, because of the near zero loadings of the variables on those components. As a high canonical correlation does not tell us

anything about the communality of two sets of variables, it is as such an analytical tool which is hard to interpret. However, Stewart and Love (1968) introduced the redundancy index, which is the mean variance of the variables of one set that is explained by a canonical variate of the other set. This index was computed for all CCA pairs and its values for CCA1 for precipitation given the predictors shown in Figure 2-3 are as follows: 0.53 for mean sea level pressure, 0.62 for 2-m air temperature, 0.78 for 850-hPa meridional wind, 0.46 for 850-hPa relative humidity, 0.41 for 10-m zonal wind, 0.38 for 500-hPa specific humidity, 0.76 for 500-hPa vertical velocity, and 0.73 for precipitable water. High redundancy suggests, perhaps, high ability of the covariates to predict precipitation.

It is not only important to understand the spatial dependence but also the pair-to-pair temporal dependence between precipitation and significant large-scale climate predictors. Figure 2-4 shows the canonical expansion coefficients/vectors of the first two CCA spatial patterns of air temperature, relative humidity, specific humidity and precipitable water. It is evident that the covariates not only show strong spatial dependence but also that similar relationships prevail on different temporal scales. However, to form feature vectors, only spatial CCA modes were used.

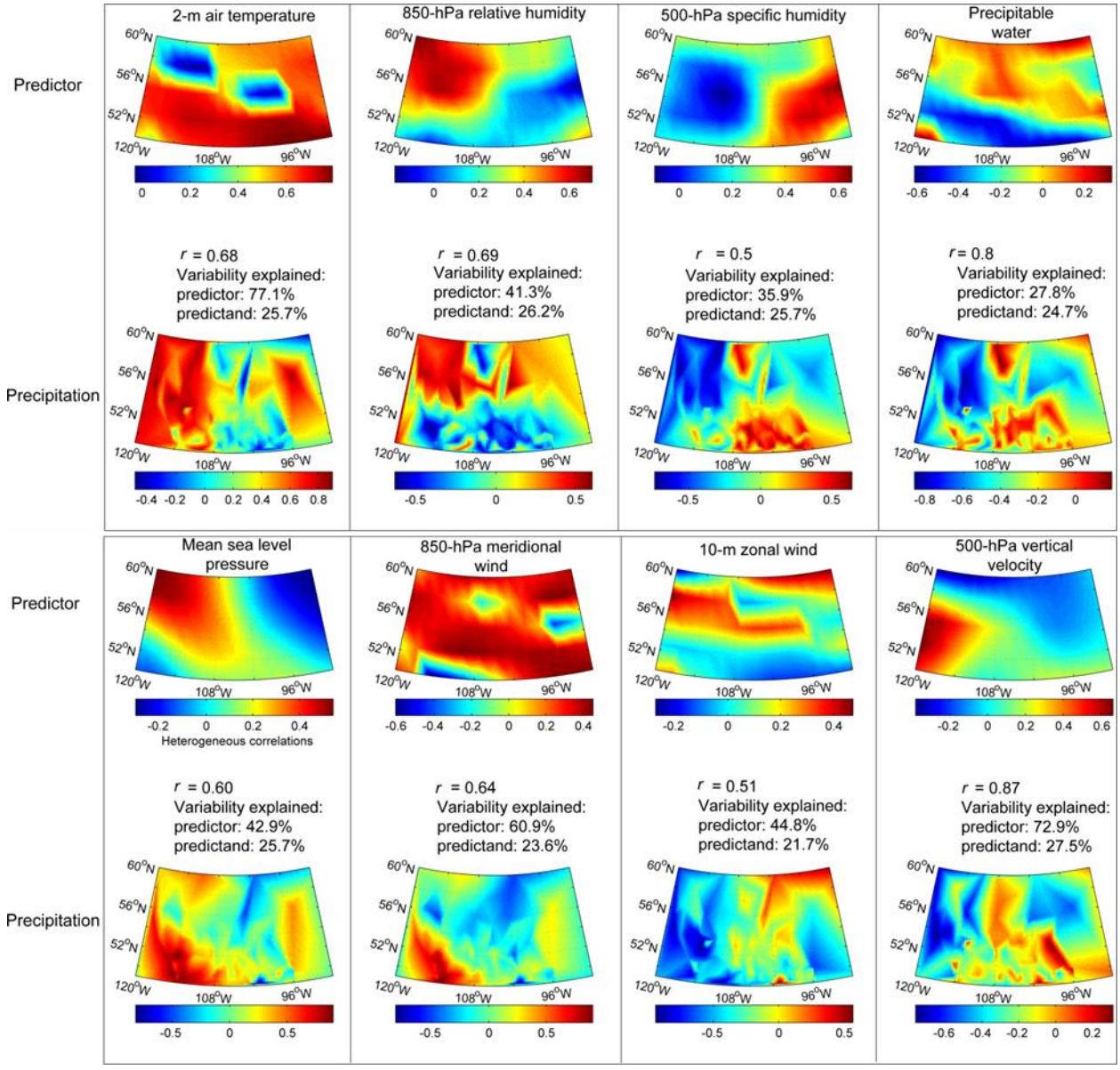

Figure 2-3: Spatial patterns of first CCA (CCA1) with heterogeneous correlations between monthly precipitation and statistically significant covariates. CCA1 pairs are shown column-wise; for example, CCA1 explains 77.1% and 25.7% of the variability in 2-m air temperature and precipitation, respectively, with significant positive canonical correlation, $r = 0.68$

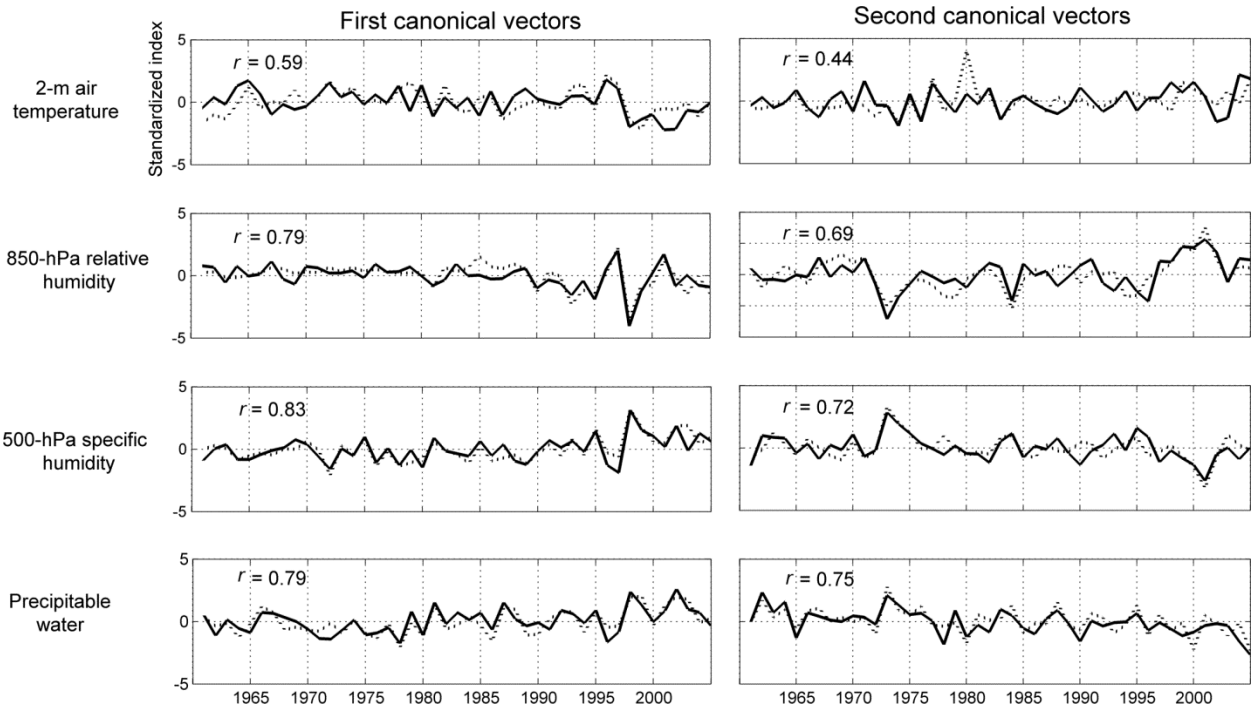

Figure 2-4: First and second canonical vectors and correlation coefficients of selected NCEP based atmospheric covariates: air temperature, relative humidity, specific humidity and precipitable water. The canonical vectors of both fields are positively correlated for the time period considered. Dashed and solid lines represent covariates and precipitation, respectively

The results of the CCA (Figure 2-5) showed that the geographical site attributes are spatially correlated with monthly precipitation. Two CCA spatial patterns which together preserved 98.56% of the variance were used as additional attributes to form input feature vectors for the FCM clustering algorithm. A high positive correlation can be seen between these fields. In summary, 12 precipitation-sensitive attributes (i.e. indices of PDO and PNA, eight NCEP-based atmospheric covariates and two CCA modes computed from site attributes) were considered for forming input feature vectors for the FCM clustering algorithm.

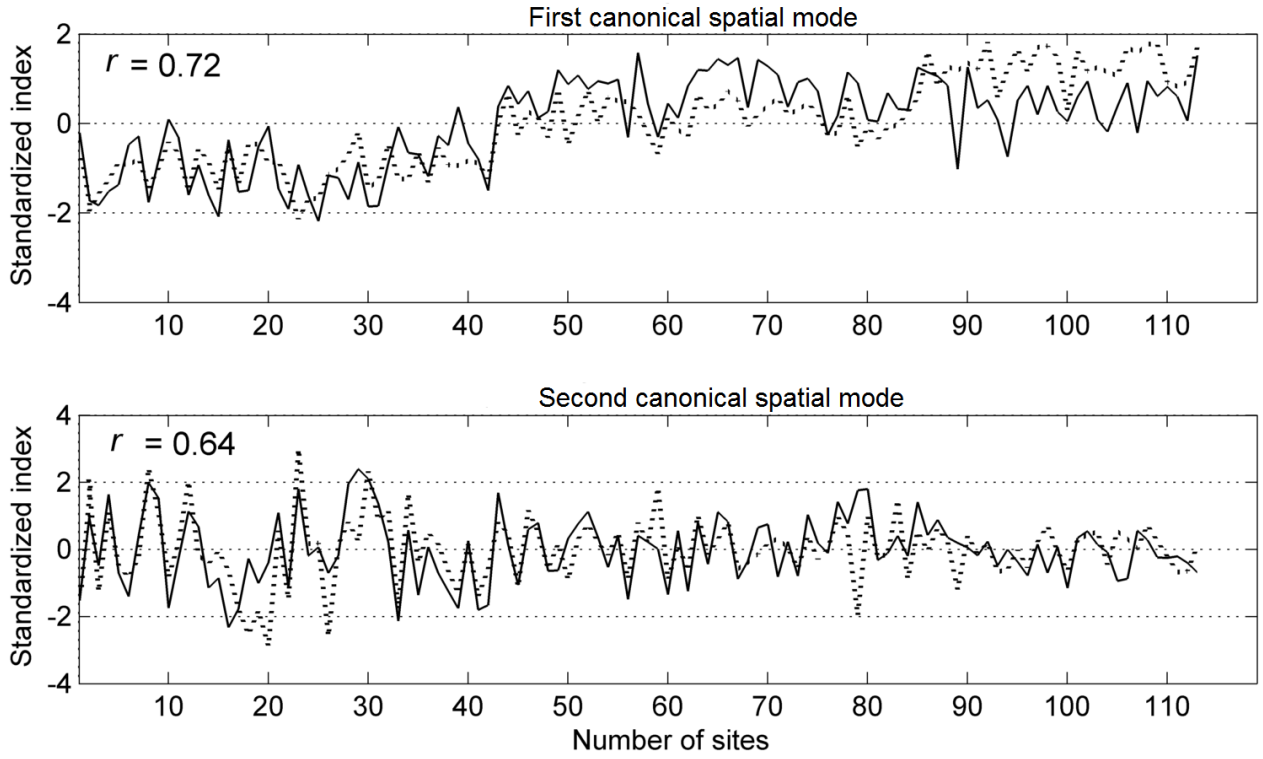

Figure 2-5: First and second canonical spatial modes derived from monthly precipitation and geographical site attributes. Precipitation amount at each site is positively correlated with latitude, longitude and elevation. Dashed and solid lines represent geographic site attributes and precipitation, respectively

2.4.2 Delineation of Climatic Regions using the FCM Clustering Algorithm

Based on the results of CCA discussed above, eight significant surface and upper level atmospheric covariates for 65 grid points (i.e. $8 \times 65 = 520$ features) were considered for additional analyses. The 520 features were dimensionally reduced using the PCA, which resulted in eight PCs that explained 97.83% of the variance. These eight PCs, along with indices of PDO and PNA and two CCA modes derived from geographical site attributes form input feature vectors for the FCM clustering algorithm.

Bezdek and Pal (1995) remarked that the FCM algorithm provides better performance for the weight exponent (fuzzifier) μ in the 1.5–2.5 range. Results of a preliminary analysis indicated that four soft clusters (c = 4) could be delineated using a combination of the Xie-Beni

cluster validity index, and $\mu = 2$. A larger value of μ implies that the objects whose characteristics are most dissimilar to the average characteristics of the centroids are less penalized (Rao and Srinivas, 2008). These preliminary regions were evaluated for homogeneity using statistics computed from observed annual precipitation; this resulted in just one region being homogeneous. Following this, the sensitivity of the results from the FCM algorithm to variation in the value of the fuzzifier was tested by varying μ from 1.2 to 2.5 with an increment of 0.1. Based on these diagnostics, it was found that fuzziness in the resulting clusters increases for μ greater than 1.6, which is directly reflected in the values of the Xie-Beni index that increases significantly beyond μ equals to 1.6, indicating a drop in the quality of the resulting clusters. Thus, the clusters given by the FCM algorithm with $\mu = 1.6$ were selected as optimal partitions. The resulting soft clusters are displayed over the 65-grid spatial domain in Figure 2-6. After identifying five fuzzy clusters, each site was assigned a membership following the iterative procedure described in Bezdek (1981). Figure 2-6 shows cluster centroids and the degree of belongingness of the *k*th feature vector in the *i*th fuzzy cluster.

The centroids are centered over southeast of the Canadian Prairie Provinces (Region A), the northern parts (Region B), the south-central part (Region C), the central west part (Region D) and the foothills of the Rocky Mountains to the southwest (Region E). Subsequently, the soft regions are evaluated independently for homogeneity using statistics computed from observed precipitation records.

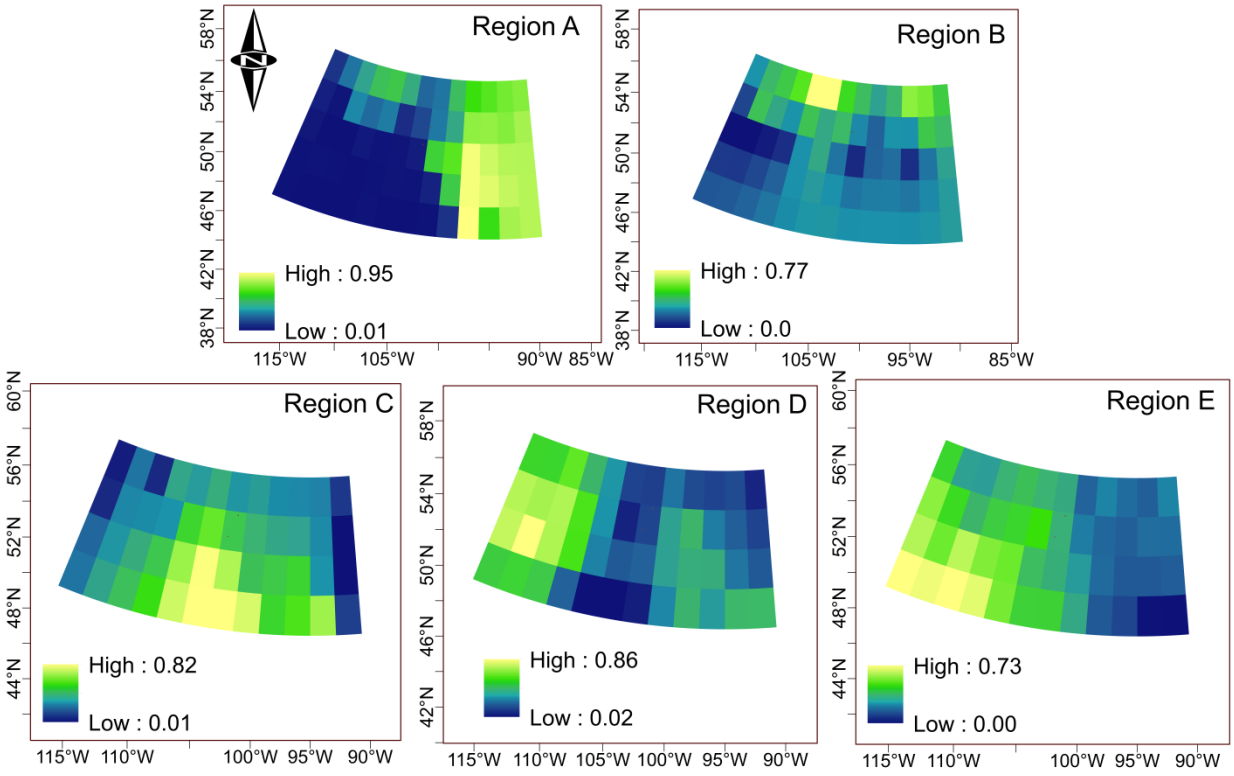

Figure 2-6: Non-defuzzified soft regions delimited by the FCM clustering algorithm using the significant atmospheric covariates, teleconnection indices and geographical site attributes. Yellow color indicates higher similarity of attributes at surrounding grids to the cluster centroids

2.4.3 Validation of Homogeneity of Soft Climatic Regions

The homogeneity of clusters corresponding to optimal partitions (obtained with $\mu = 1.6$), formed without defuzzification, was tested using heterogeneity measures of Hosking and Wallis (1997) by computing statistics from annual and seasonal precipitation totals, and seasonal extremes. Firstly, the number of sites (out of 120) falling in each of the five soft regions was determined. Subsequently, the discordancy statistic *D* was computed for all sites for each of the five regions separately. Sites with $D \geq 3$ in a region were considered discordant. According to this assessment, no sites were found to be discordant. The number of sites for each region is given in Table 2-2. Regionally averaged values of L-CV, L-skewness and L-kurtosis are also provided in this table. The values of L-CV suggested that the monthly precipitation is relatively

more variable in Region E compared to other regions. Similarly, the distributions of monthly precipitation in Region A and Region C show relatively higher degree of symmetry compared to other regions. The values of the heterogeneity measures H_k for all regions are summarized in Table 2-3. According to the results shown in this table, all five regions can be judged as "acceptably homogeneous" at the monthly timescale. Spatial distribution of sites among the homogeneous soft regions is shown in Figure 2-7. It is evident from this Figure that most sites with shared memberships are distributed around boundaries of the soft regions.

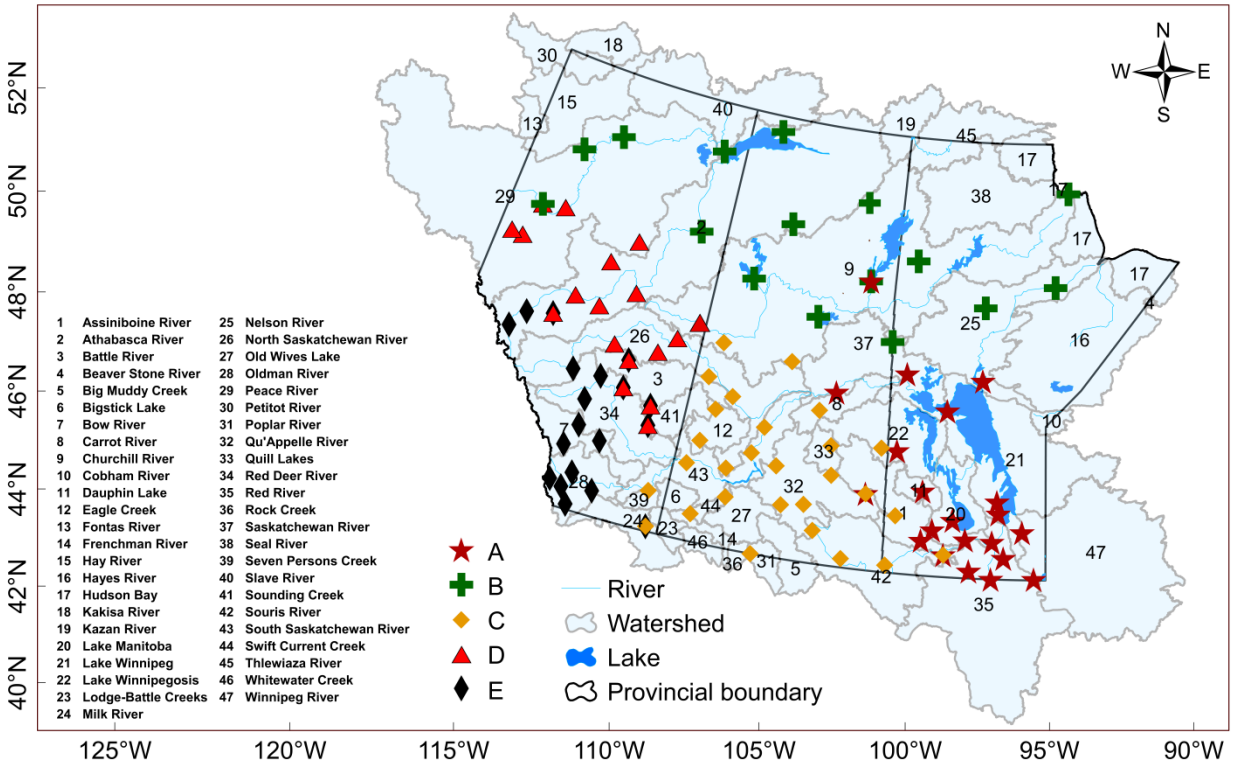

Figure 2-7: Delineated homogeneous precipitation regions (A, B, C, D and E), along with spatial distribution of precipitation observation stations

After validation of the homogeneous regions at the monthly time scale, statistical homogeneity of fuzzy regions was evaluated at the seasonal timescale using statistics derived from

Table 2-2: Regional L-moments (L-CV, L-skewness and L-kurtosis) of monthly precipitation and number of sites associated with each of the five soft regions

Region	L-CV	L-skewness	L-kurtosis	Number of sites
A	0.310	0.09	0.15	23
B	0.294	0.21	0.56	16
C	0.256	0.16	0.38	41
D	0.121	0.19	0.12	19
E	0.432	0.28	0.30	21

Table 2-3: Values of the heterogeneity measures H_k for each of the five regions for the case of monthly precipitation totals

Region	H_1	H_2	H_3
A	0.890	0.640	0.817
B	0.243	0.432	-0.177
C	0.725	0.401	0.026
D	-0.144	-0.820	-0.717
E	0.403	0.833	0.638

winter and summer precipitation totals. Results of this evaluation are summarized in Table 2-4 and the spatial distribution of sites among the soft regions is shown in Figure 2-8. Based on the discordancy measure *D* for winter precipitation, no site was judged as discordant. A similar assessment for summer precipitation indicates that just one site in Region D could be marginally discordant. For both seasons, all regions can be judged as "acceptably homogeneous" (Table 2-4).

Table 2-4: Values of the heterogeneity measures H_k for each of the five regions for the case of winter and summer seasonal precipitation totals

Region	Winter			Summer		
	H_1	H_2	H_3	H_1	H_2	H_3
A	0.67	0.83	-0.77	0.84	0.89	0.34
B	0.91	0.40	0.80	0.97	0.68	0.27
C	0.92	-0.26	0.52	0.89	0.98	0.60
D	0.52	0.49	0.61	-0.34	-0.14	1.02
E	0.51	-0.74	0.05	0.29	0.63	-0.41

In addition, no significant differences were noted when a similar validation (see Table C1 in Appendix C) was performed after defuzzification of the soft clusters following suggestions from

Jonathan Hosking (personal communication, June 23, 2013). However, H_k values are found slightly better in the case of soft regions without defuzzification.

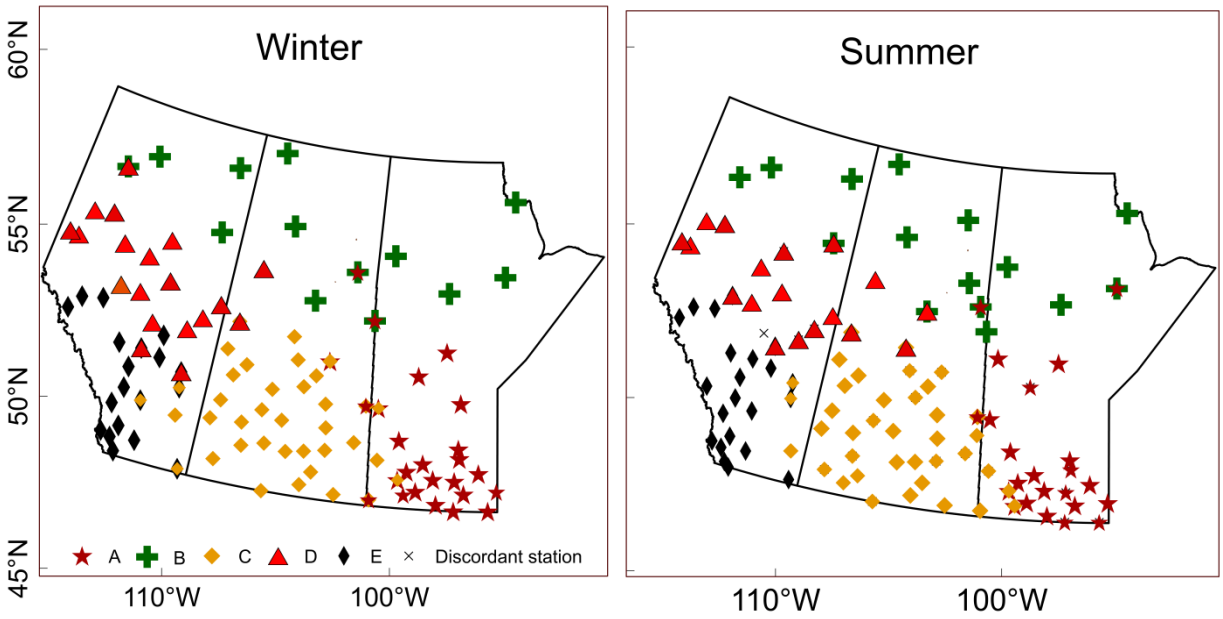

Figure 2-8: Validation of the FCM clustering algorithm based homogeneous precipitation regions (A, B, C, D and E) using statistics derived from winter and summer seasonal precipitation totals

The validity of the homogeneity of the soft regions was also evaluated using statistics computed from cold and warm season precipitation extremes. Prior to this evaluation, the soft clusters were defuzzified. Values of the heterogeneity measures H_k for each of the five defuzzified regions for the case of cold and warm season precipitation extremes are shown in Table 2-5. None of the stations was found grossly discordant from the group as a whole. Close examination of the results reported in Table 2-5 indicated that H_2 is exceeded in Region C for 1-day winter extremes. Aside from this, the regions can be declared as "acceptably homogeneous", making them suitable for RFA. It is important to mention that on two different time scales, cluster centroids (for non-defuzzified regions) remain stable but minor instabilities near the boundaries are seen. Stations at boundaries of clusters tend to have partial or distributed

memberships among two or more clusters. However, this did not impact the overall results of homogeneity validation presented above.

Table 2-5: Values of the heterogeneity measure H_k for each of the five defuzzified regions for the case of winter and summer seasonal 1-, 3- and 5-day precipitation extremes

Region	1-day			3-day			5-day		
	Winter								
	H_1	H_2	H_3	H_1	H_2	H_3	H_1	H_2	H_3
A	0.56	0.38	0.28	0.58	-0.83	-0.54	0.17	-0.15	-0.25
B	0.94	-0.17	0.28	0.41	-0.58	0.50	0.24	0.59	0.86
C	0.99	1.26	0.54	0.42	0.27	0.50	0.10	0.92	0.24
D	0.59	0.35	0.69	0.01	0.32	0.76	0.54	0.91	0.89
E	0.10	0.69	0.45	0.80	0.13	0.56	0.27	0.10	0.30
	Summer								
A	0.24	0.69	-0.02	-0.21	0.19	0.47	0.66	0.26	0.42
B	-0.15	0.19	0.76	0.55	0.33	0.06	0.71	-0.25	0.08
C	0.72	0.03	0.01	0.64	0.51	0.90	0.04	0.80	0.14
D	0.95	0.46	0.09	0.80	0.48	0.04	0.98	0.73	0.72
E	0.22	-0.97	-0.40	0.40	0.96	0.08	0.91	0.50	0.32

The foregoing analyses have shown that large-scale atmospheric covariates, indices of teleconnection patterns and geographic site attributes can be used to define homogeneous precipitation regions over the three Prairie Provinces of Canada. It is recommended that future work should investigate other attributes such as the physiographic characteristics of watersheds and indices of temperature extremes. The methodology proposed here differs considerably from other similar approaches suggested by Comrie and Glenn (1998) and Ouarda et al. (2001), who respectively used PCA and CCA to identify homogeneous regions. Comrie and Glenn (1998) illustrated applicability of their regionalization approach via an analysis of relationships between monsoon precipitation variability and 500 mb pressure heights. However, such large scale circulation is instead used in this study as a candidate predictor in the clustering routine. The approach of Ouarda et al. (2001) is based on the concept of hydrological neighborhood and it generally leads to noncontiguous regions. Furthermore, an independent validation of the

delineated regions is not possible in these two approaches, but it is possible in the approach proposed in this study.

2.4.4 Regional Frequency Analysis of Seasonal Precipitation Extremes

2.4.4.1 Tests for stationarity and autocorrelation

The results of the original Mann-Kendall (MK) test reveal that for 1-day precipitation extremes altogether seven stations showed a statistically significant trend at the 5% significance level. However, when the modified MK test was applied to the same dataset, no significant trends were noted. Thus, the presence of non-stationarity appears to be due to autocorrelation in time series of precipitation extremes. By examining the lag-1 autocorrelation in samples of seasonal precipitation extremes, it is found that 13 stations showed significant serial dependence at 5% significance level in one of the three time series of extremes, i.e. 1-, 3- and 5-day extremes. It is important to note that the significance of lag-1 autocorrelations disappeared at the 10% significance level. Therefore, in the presence of only a weak non-stationary signal and moderate serial dependence in a handful of time series, it is assumed that the entire dataset was suitable for RFA of seasonal precipitation extremes.

2.4.4.2 Choice of a regional frequency distribution and estimation of regional growth curves

Six commonly used distributions (i.e. GLO, GNO, GPA, PE3, GEV and Wak) were evaluated for modeling regional growth curves for seasonal single- and multiday precipitation extremes. Based on the LMR diagrams and *Z*-statistic, multiple distributions could be possible candidates. In summary, from the six considered distributions, single- and multiday seasonal precipitation extremes for most regions can be modeled by using the GNO, GEV and PE3 distributions. After a careful scrutiny, the GEV, PE3, GEV, PE3, and GNO were chosen as the possible best-fit candidates for cold season precipitation extremes for Region A, Region B,

Region C, Region D and Region E, respectively. Similarly, for Region A, Region B, Region C, Region D and Region E, the GEV, GNO, GEV, GNO and GNO, respectively, were selected as the

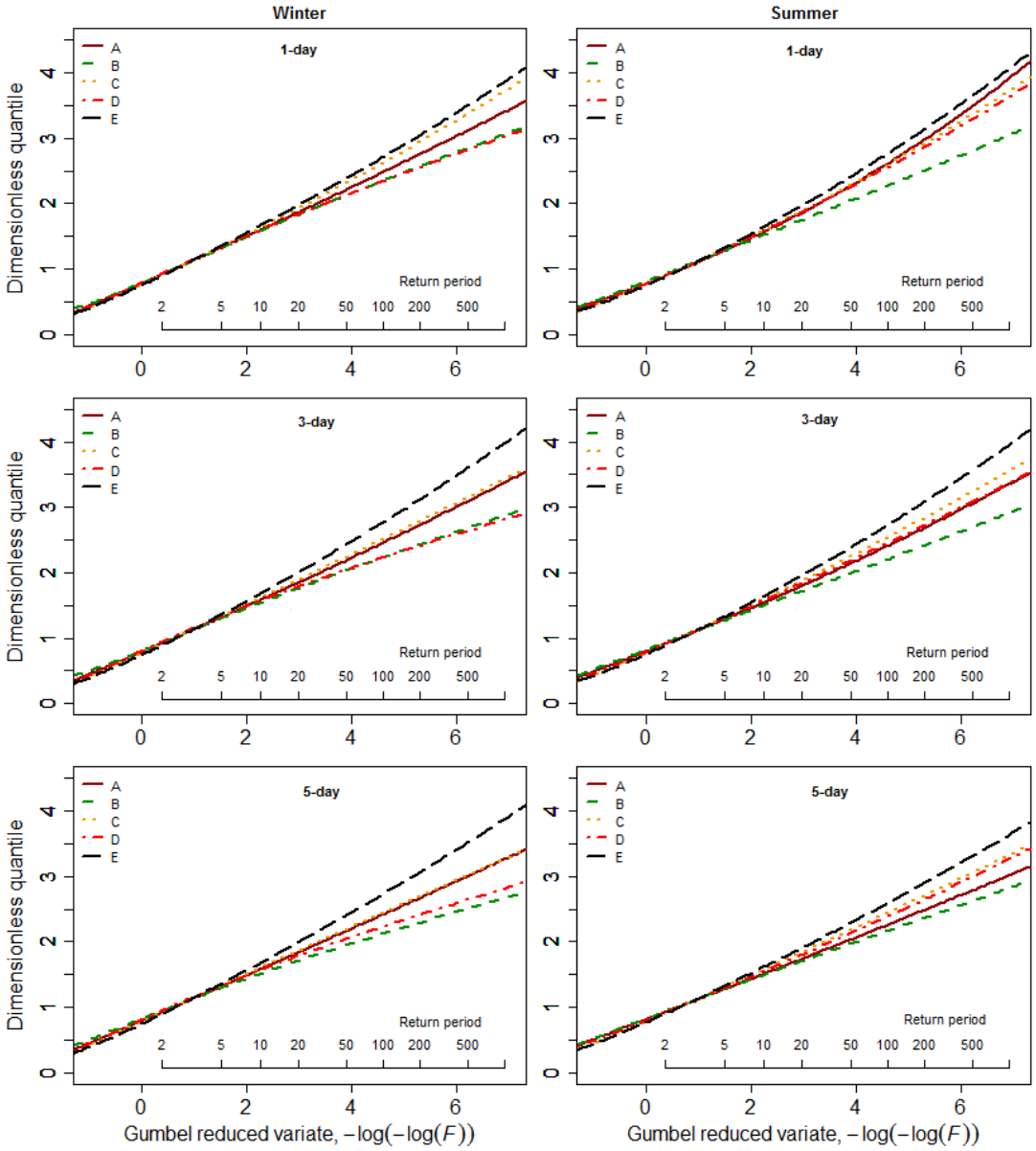

Figure 2-9: Regional growth curves of 1-, 3-, and 5-day winter and summer season precipitation extremes. These curves are based on the selected regional frequency distribution, discussed in the main text. Inner scale along the x-axis represents return period and *F* is the cumulative probability of non-exceedance

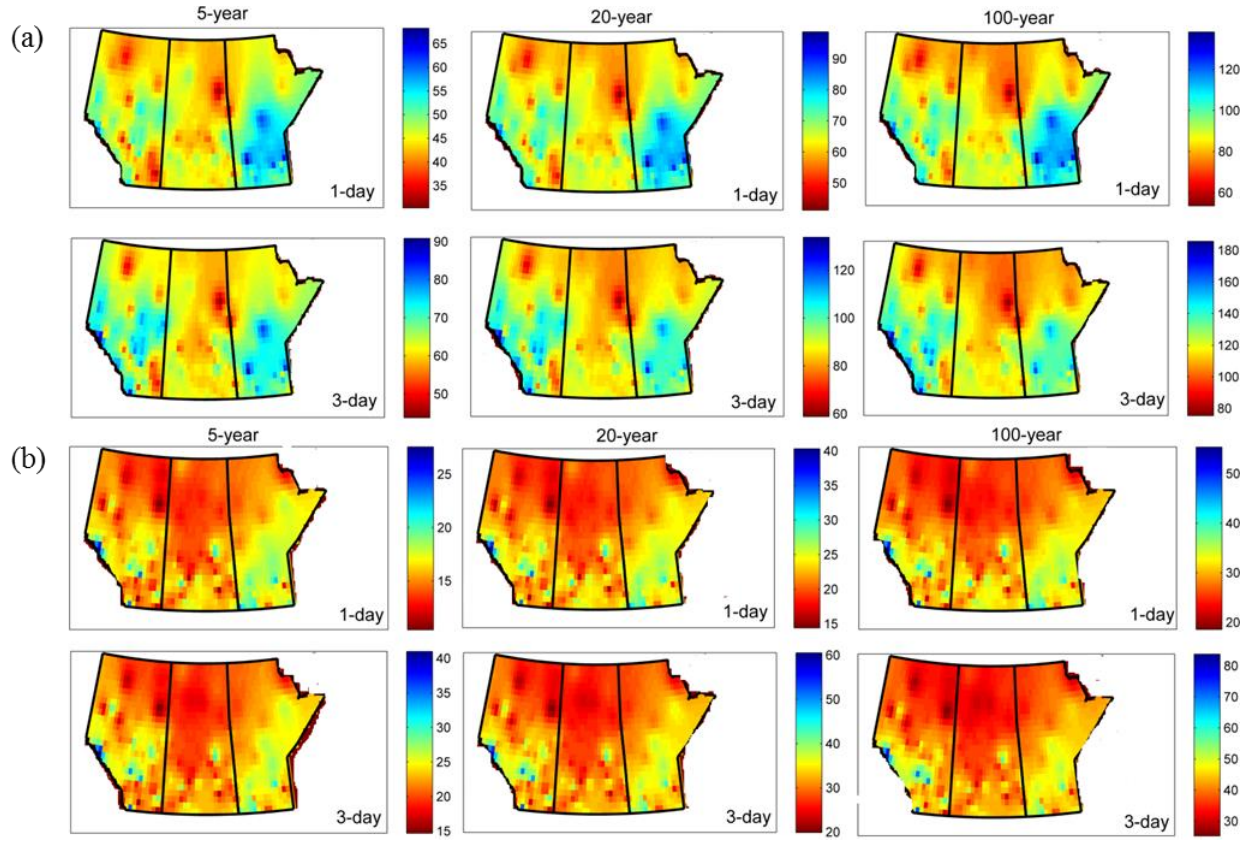

Figure 2-10: Spatial distribution of 5- (column 1), 20- (column 2) and 100-year (column 3) return values of 1- and 3-day precipitation extremes (in mm) for (a) summer and (b) winter seasons. Different scales are used across return values and duration of extremes to display spatial variability

best-fit distributions for warm season precipitation extremes. The best-fit distribution identified for each homogeneous region was used to estimate the regional growth curve and the associated growth factors corresponding to various return periods selected. These growth curves are shown in Figure 2-9. It is evident from this Figure that the Region E is associated with the highest values of growth factors for all single- and multi-day precipitation extremes for both winter and summer seasons. Return values of single- and multi-day precipitation extremes for all stations belonging to a given region were calculated by multiplying the regional growth factors with the at-site mean values. From the at-site return values, spatial patterns for the entire study domain were developed using the ordinary kriging-based interpolation method of Oliver (1990). Results

for 5-, 20- and 100-year return values of 1- and 3-day precipitation extremes are shown in Figure 2-10 for summer and winter seasons. Results for 5-day extremes (see Figure C1in Appendix C) are generally similar to those of 3-day extremes.

In general, the spatial patterns indicate that precipitation return values tend to be of higher magnitude for the western and southeastern watersheds in the study area, both in winter (Figure C1in Appendix C) and summer. On the contrary, watersheds in the Prairies and the northern regions tend to experience return values of smaller magnitude. On the basis of a relative comparison of return values of 1-, 3- and 5-day extremes, areas with higher return values of single-day extremes are also associated with higher return values of multiday sequences. In winter, 1-day precipitation extremes in the west exhibit a tendency towards a decrease in the spatial extent compared to multiday events, while the multiday extremes tend to reduce in the spatial extent relative to 1-day events in the eastern parts of the study area. Using the 100-year return values of 3-day summer extremes as an example, the smallest values are found over the Churchill River, Athabasca River, Seal River, Lodge-Battle Creeks and South Saskatchewan River watersheds (see Figure 2-6 for correspondence). The spatial extent of the magnitude of single- and multiday extremes is generally larger in summer compared to winter.

2.5 Summary and Conclusions

A new approach for regionalization of precipitation climates was proposed in this study for the Canadian Prairie Provinces of Alberta, Saskatchewan and Manitoba. This part of Canada offers significant scientific challenges due to its highly variable hydro-climate and therefore demands quite complex methods to be explored. Multivariate approaches like PCA and CCA were utilized to screen large-scale atmospheric covariates derived from NCEP reanalysis, indices of teleconnection patterns and geographic site attributes in order to form feature vectors for the FCM clustering algorithm for delineating homogeneous precipitation regions. Based on the

identified homogeneous regions, various return values of cold and warm season precipitation extremes, derived within the framework of RFA, were spatially mapped across the entire study domain. Such information about precipitation extremes is often required for water resources development and management related activities. Based on the analyses presented in this study, the following main conclusions can be drawn:

(1) For determining possible large-scale forcing of the observed variability in precipitation, PCA applied to monthly precipitation totals resulted in 8 PCs that explained about two-thirds of the observed variability in precipitation. From the set of five teleconnection patterns, the indices of PDO and PNA were found to be significantly correlated with regional modes of precipitation variability for the entire study area.

(2) Statistically significant coherent heterogeneous canonical spatial and temporal patterns between NCEP-based atmospheric covariates and the corresponding patterns in the precipitation field were seen and therefore, it can be deduced that the precipitation variability in the study area is probably linked with the variability in the large-scale atmospheric circulation at the monthly timescale.

(3) Based on the information derived from indices of teleconnection patterns, large scale atmospheric covariates and geographic site attributes used in the FCM clustering algorithm, the Canadian Prairie Provinces can be sub-divided into five homogeneous climatic regions: Region A–southeast of the Canadian Prairies; Region B–the northern region; Region C–the south-central region; Region D–the central west; Region E–the foothills of the Rocky Mountains to the southwest.

(4) An independent validation based on the L-moments approach (Hosking and Wallis 1997) and statistics derived from monthly and seasonal precipitation totals, and single- and multiday seasonal precipitation extremes from a network of observation stations located across the

study area confirmed statistical homogeneity of the identified regions. Consequently, the identified regions can be judged as statistically as well as climatologically homogeneous. This is a significant step forward, since according to the conventional approach, only statistical homogeneity can be assured.

(5) Based on the identified homogeneous regions, various return values of cold and warm season single- and multiday precipitation extremes were developed within the framework of RFA approach and mapped across the study domain. Spatial maps of various return values indicated that precipitation extremes tend to be of higher magnitude over the western and southeastern watersheds, both in winter and summer, while for the watersheds in the Prairies and the northern regions they tend to be of smaller magnitude. On the basis of comparative evaluation of return values of single- and multiday precipitation extremes, it can be concluded that the areas which receive relatively larger single-day precipitation are also the same which receive larger multiday sequences.

Finally, the proposed framework offers promising opportunities for studying the spatial and temporal dynamics of hydro-meteorological variables. To the best of our understanding, the present study is the first attempt to perform a systematic RFA of single- and multiday precipitation extremes over the Canadian Prairie Provinces for regions which are statistically and climatologically homogeneous. The proposed methodological framework constitutes a major contribution towards a better understanding of the linkages between precipitation characteristics and large-scale atmospheric circulations in a regional perspective. The results of this study form a basis for developing a multisite precipitation downscaling approach in future for the three Prairie Provinces of Canada.

Acknowledgements

The financial support from the Global Institute for Water Security and School of Environment and Sustainability is gratefully acknowledged. Thanks are due to Eva Mekis from Environment Canada for providing access to adjusted precipitation dataset used in this study. We would like to express our thanks to Jonathan Hosking for his advice on the defuzzification of the fuzzy soft clusters prior to applying the regional frequency analysis algorithm. Thanks are also due to Sun Chun for useful comments which helped improve the analyses presented in this paper. The authors further thank three anonymous reviewers for their helpful comments on the manuscript.

References

Adamowski K, Alila Y, Pilon JP (1996), Regional rainfall distribution for Canada, *Atmospheric Research,* 42, 75–88.

Alila Y (1999), A hierarchical approach for the regionalization of precipitation annual maxima in Canada, *Journal of Geophysical Research,* 104, 31645–31655.

Baeriswyl PA, Rebetez M (1997), Regionalization of precipitation in Switzerland by means of principal component analysis, *Theoretical and Applied Climatology,* 58, 31–41.

Basalirwa CPK (1995), Delineation of Uganda into climatological rainfall zones using the method of principal component analysis, *International Journal of Climatology,* 15(10), 1161–1177.

Bezdek JC (1981), Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York

Bezdek JC, Pal NR (1995), Two soft relatives of learning vector quantization, *Neural Networks* 8(5), 729–743.

Bobée B, Rasmussen P (1995), Recent advances in flood frequency analysis. US National Report to International Union of Geodesy and Geophysics 1991–1994, *Reviews in Geophysics* 33, 1111–1116.

Bonsal BR, Aider R, Gachon P, Lapp S (2012), An assessment of Canadian prairie drought: past, present, and future, *Climate Dynamics,* 41, 501–516.

Bonsal BR, Lawford RG (1999), Teleconnections between El Niño and La Niña events and summer extended dry spells on the Canadian prairies, International *Journal of Climatology*, 19, 1445–1458.

Borchert JA (1950), The Climate of the Central North American grassland, *Annals of the Association of American geographers,* 40, 1–39.

Brown RD, Goodison BE (1996), Interannual variability in reconstructed Canadian snow cover 1915-1992, *Journal of Climate*, 9, 1299–1318.

Bürger K (1958), Zur Klimatologie der Grosswetterlagen, Ber. Dtsch. Wetterdienstes 45, 6, Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main, Germany

Burn DH (1988), Delineation of groups for regional flood frequency analysis, *Journal of Hydrology,* 104, 345–361.

Burn DH (1989), Cluster analysis as applied to regional flood frequency analysis, *Journal of Water Resources Planning and Management* 115, 567–82.

Burn DH (1990a), An appraisal of the region of influence approach to flood frequency analysis, *Hydrological Sciences Journal* 35(2), 149–165.

Burn DH (1990b), Evaluation of regional flood frequency analysis with a region of influence approach, *Water Resources Research* 26(10), 2257–2265.

Cavadias GS (1989), Regional flood estimation by canonical correlation, paper presented to the 1989 Annual Conference of the Canadian Society for Civil Engineering, St-John's, Newfoundland

Cavadias GS (1990), The canonical correlation approach to regional flood estimation, Regionalization in hydrology, In Proceedings of the Ljubljana Symposium, IAHS Publication number 191, IAHS, Wallingford

Chandler RE, Wheater HS (2002), Analysis of rainfall variability using generalized linear models-a case study from the West of Ireland, *Water Resources Research* 38(10), 10-1–10-11.

Comrie AC, Glenn EC (1998) Principal components-based regionalization of precipitation regimes across the southwest United States and northern Mexico, with an application to monsoon precipitation variability, *Climate Research,* 10, 201-215.

Cosmo SN, Xu CY, Tallaksen LM, Alemaw B, Chirwa T (2011), Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and L-moments approaches, *Stochastic Environmental Research and Risk Assessment,* 25, 939-955.

Douglas EM, Vogel RM, Kroll CN (2000), Trends in floods and low flows in the United States: impact of spatial correlation, Journal of Hydrology, 240, 90–105.

Dunn JC (1974), A fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters, *Journal of Cybernetics*, 3, 32–57.

Ehrendorfer M (1987), A regionalization of Austria's precipitation climate using principal component analysis, *Journal of Climatology*, 7(1), 71–89.

Fowler HJ, Kilsby CG, O'Connell PE (2000), A stochastic rainfall model for the assessment of regional water resource systems under changed climatic conditions, *Hydrology and Earth System Science*, 4, 263–282.

Gadgil S, Yadumani, Joshi NV (1993), Coherent rainfall zones of the Indian region, *International Journal of Climatology*, 13(5), 547–566.

Ge Y, Gong G (2009), North American snow depth and climate teleconnection patterns. *Journal of Climate*, 22, 217–233.

Ghatak D, Gong G, Frei A (2010), North American temperature, snowfall, and snow-depth response to winter climate modes, *Journal of Climate*, 23, 2320–2332.

Gottschalk L (1985), Hydrologic regionalization of Sweden, *Hydrological Science Journal*, 30, 65–83.

Graham S (1988), Precipitation: Process and Analysis. John Wiley and Sons Ltd., Baffins Lane, Chichester, UK

Guttman NB (1993), The use of L-Moments in the determination of regional precipitation climates, *Journal of Climate*, 6(12), 2309–2325.

Guttman NB, Hosking JRM, Wallis JR (1993), Regional precipitation quantile values for the continental United States computed from L-Moments, *Journal of Climate*, 6, 2326–2340.

Hall MJ, Minns AW (1999), The classification of hydrologically homogeneous regions, *Hydrological Sciences Journal*, 44(5), 693–704.

Hamed KH, Rao AR (1998), A modified Mann Kendall trend test for autocorrelated data, *Journal of Hydrology*, 204, 182-196.

Hosking JRM, Wallis JR (1997), Regional Frequency Analysis: An Approach Based on L-moments. Cambridge University Press, New York, USA

Kalnay E et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bulletin of American Meteorological Society*, 77(3), 437 – 471.

Kannan S, Ghosh S (2011), Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output, *Stochastic Environmental Research and Risk Assessment*, 25, 457-474.

Kendall MG (1975), Rank Correlation Methods. Griffin, London

Khaliq MN, Gachon P (2010), Pacific Decadal Oscillation climate variability and temporal pattern of winter flows in north-western North America, *Journal of Hydrometeorology*, 11, 917–933.

Krzanowski WJ (1988), Principles of Multivariate Analysis: A User's Perspective. Oxford University Press, New York

Lamb HH (1972), British Isles weather types and a register of the daily sequence of circulation patterns, 1861–1971. *Geophysical Memoir* 116HMSO, London

Lamb HH (1977), Climate: Present, Past and Future: Climatic history and the future, Volume 2, Methuen, London

Lin GF, Chen LH (2006), Identification of homogeneous regions for regional frequency analysis using the self-organizing map, *Journal of Hydrology*, 324, 1–9.

Loikith PC, Broccoli AJ (2012), Characteristics of observed atmospheric circulation patterns associated with temperature extremes over North America, *Journal of Climate*, 25, 7266–7281.

Maheras P, Tolika K, Anagnostopoulou C, Vafiadis M, Patrikas I, Flocas H (2004), On the relationships between circulation types and changes in rainfall variability in Greece, *International Journal of Climatology*, 24(13), 1695–1712.

Mann HB (1945), Nonparametric tests against trend, *Econometrica*, 13, 245–259.

Mantua NJ, Hare SR (2002), The Pacific Decadal Oscillation, *Journal of Oceanography* 58, 35–44.

McQueen J (1967), Some methods for classification and analysis of multivariate observations, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1, University of California Press, Berkeley, pp. 281–297

McQuitty LL (1957), Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies, *Educational and Psychological Measurement*, 17, 207–229.

Mekis É, Vincent LA (2011), An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada, *Atmosphere-Ocean* 49(2), 163–177.

Mosley MP (1990), Delimitation of New Zealand hydrologic regions, *Journal of Hydrology*, 49(1-2), 173–192.

Ouarda TBMJ, Girard C, Cavadias GS, Bobée B (2001), Regional flood frequency estimation with canonical correlation analysis, *Journal of Hydrology*, 254, 157–173.

Oliver MA (1990), Kriging: A Method of Interpolation for Geographical Information Systems, *International Journal of Geographic Information Systems*, 4(3), 313–332.

Perez-Valdivia C, Sauchyn D, Vanstone J (2012), Groundwater levels and teleconnection patterns in the Canadian Prairies, *Water Resources Research*, 48 W07516, doi:10.1029/2011WR010930.

Philips D (1990), Climates of Canada. Canadian Government Publishing Ottawa, Ontario

Preisendorfer R, Barnett TP (1987), Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by Canonical Correlation Analysis, *American Meteorological Society*, 115, 1825–1850.

Preisendorfer RW, Mobley CD (1988), Principal Component Analysis in Meteorology and Oceanography. Elsevier Science Publishing Company, Inc., Amsterdam

Rao AR, Srinivas VV (2006b), Regionalization of watersheds by fuzzy cluster analysis, *Journal of Hydrology*, 318, 57 – 79.

Rao AR, Srinivas VV (2008), Regionalization of Watersheds – An Approach Based on Cluster Analysis. Springer Publishers, Germany

Reed DW, Jakob D, Robson AJ (1999), Selecting a pooling group. In: Robson AJ, Reed DW (ed) Statistical Procedures for Flood Frequency Estimation. Flood Estimation Handbook, volume 3, Institute of Hydrology, Wallingford, UK

Rianna M, Ridolfi E, Lorino L, Alfonso L, Montesarchio V, Di Baldassarre G, Russo F, Napolitano F (2012), Definition of homogeneous regions through entropy theory. Third STAHY International Workshop on statistical methods for hydrology and water resources management, Tunis, Tunisia

Richman MB (1986), Rotation of principal components, *International Journal of Climatology*, 6, 293-335.

Ripley EA (1992), Grassland Climate, in Natural grasslands: Introduction and Western Hemisphere, Ecosystems of the World, volume 8A, edited by Coupland RT pp. 7–24, Elsevier, Amsterdam.

Romolo L, Prowse TD, Blair D, Bonsal BR, Martz LW (2006a), The synoptic climate controls on hydrology in the upper reaches of the Peace River Basin. Part I: snow accumulation, *Hydrological Processes,* 20, 4097–4111.

Ropelewski CF, Halpert MS (1986), North American Precipitation and Temperature Patterns Associated with the El Niño/Southern Oscillation (ENSO), *Monthly Weather Review,* 114, 2352–2362.

Rossi F, Villani P (1994), Regional flood estimation methods. In: Rossi G, Harmancioglu NB, Yevjevich V (ed) Coping with Floods. Kluwer Academic Publishers, Dordrecht, pp., 135–169

Satyanarayana P, Srinivas VV (2011), Regionalization of precipitation in data sparse areas using large scale atmospheric variables – A fuzzy clustering approach, *Journal of Hydrology,* 405, 462–473.

Satyanarayana P, Srinivas VV (2008)m Regional frequency analysis of precipitation using large-scale atmospheric variables, *Journal of Geophysical Research,* 113, D24110 doi:10.1029/2008JD010412.

Shabbar A, Bonsal BR, Szeto K (2011), Atmospheric and Oceanic Variability Associated with Growing Season Droughts and Pluvials on the Canadian Prairies, *Atmosphere-Ocean* 49(4), 339–355.

Sheridan SC (2003), North American weather-type frequency and teleconnection indices, *International Journal of Climatology,* 23, 27–45.

Srinivas VV, Tripathi S, Rao AR, Govindaraju RS (2008), Regional flood frequency analysis by combined self-organizing feature map and fuzzy clustering, *Journal of Hydrology* 348, 148–166.

Stedinger JR, Vogel RM, Georgiou EF (1993), Frequency analysis of extreme events. In: Maidment DR (ed) Handbook of Hydrology, McGraw-Hill, pp. 18.1-18.66

Stewart D, Love W (1968), A general canonical correlation index. Psychological Bulletin 70: 160-163.

Szeto K, Henson W, Stewart R, Gascon G (2011), The Catastrophic June 2002 Prairie Rainstorm, *Atmosphere-Ocean,* 49,4, 380-395.

Taha BMJ, Ouarda CG, Cavadias GS, Bobée B (2001), Regional flood frequency estimation with canonical correlation analysis, *Journal of Hydrology,* 254,157-173.

Tatli H, Dalfes HN, Mentes S (2004), A statistical downscaling method for monthly total precipitation over Turkey, *International Journal of Climatology*, 24, 161–180.

Terray L, Cassou C (2000), Modes of low-frequency climate variability and their relationships with land precipitation and surface temperature: application to the Northern Hemisphere winter climate, *Stochastic Environmental Research and Risk Assessment*, 14, 339-369.

Viglione A, Laio F, Claps P (2007), A comparison of homogeneity tests for regional frequency analysis, *Water Resources Research* 43: W03428 doi:10.1029/2006WR005095.

Von Storch H, Burger G, Schnur R, Von Storch JS (1995), Principal oscillation pattern: A Review, *Journal of Climate*, 8, 377–400.

Wheater HS, Gober P (2013), Water Security in the Canadian prairies: science and management challenges, *Philosophical Transactions of the Royal Society A* (in press). doi: 10.1098/rsta.2012.0409.

Wilks SD (2011), Statistical Methods in the Atmospheric Sciences, third edition, Academic Press, San Diego, California.

Woo MK, Thorn R (2008), Analysis of cold season streamflow response to variability of climate in north-western North America, *Nordic Hydrology*, 39, 257–265.

Wu S, Yang J, Tung Y (2006), Identification and stochastic generation of representative rainfall temporal patterns in Hong Kong territory, *Environmental Research and Risk Assessment*, 20, 171–183.

Xie XL, Beni GA (1991), Validity measure for fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 841–847.

Xoplaki E, Gonzalez-Rouco JF, Luterbacher J, Wanner H (2004), Wet season Mediterranean precipitation variability: influence of large scale dynamics and trends, *Climate Dynamics*, 23, 63–78.

Zhao H, Higuchi K, Waller J, Auld H, Mote T (2013), The impacts of the PNA and NAO on annual maximum snowpack over southern Canada during 1979-2009, *International Journal of Climatology*, 33(2), 388–395.

CHAPTER 3
MULTISITE MULTIVARIATE MODELING OF DAILY PRECIPITATION AND TEMPERATURE IN THE CANADIAN PRAIRIE PROVINCES USING GENERALIZED LINEAR MODELS

Following from Chapter 1, there is a pressing need to understand in detail the possible impacts of climate change on the environment and water resources in particular. A specific area which requires immediate attention is the effect of a changing global climate on local scale hydro-meteorological variables such as precipitation, temperature, evapotranspiration, and streamflow. For example, in order to drive and evaluate hydrological models for future flood and drought risk assessment at multiple points in a large watershed or across a large geographic region, one often relies on observed weather data. But these data in most cases represent a major source of uncertainty in the modelling process. The most obvious case is when some observations are missing from available measurements. More so, the data from different sites have different lengths and in extreme situations where inputs are required at ungauged locations. As an attempt to resolve these issues missing observations are often filled in. Also, interpolation techniques have been applied to a network of weather stations to estimate required data at ungauged sites. However, interpolation often leads to smoothing and reduces natural variability in observed data.

As a potential remedy to these issues, weather generators (WGs) have emerged and are often used to generate long time-series of various weather variables suitable for risk assessment. WGs provide the means of extending data to ungauged locations, in-filling missing observations and can also serve as a computationally inexpensive tool to produce climate change scenarios. Also, these models have the ability to simulate sequences of weather variables (at any time scale) that are consistent not only with the historical climatology but also the projected changes in climate. However, most of the WGs focus on individual sites and are therefore unable to

represent the spatial structure of the observed climatic variables. Although these models can generate time series at more than one site when applied separately, the series so generated would not be spatially consistent, due to neglecting inter-station correlations. Nevertheless, for many investigations in hydrology, agriculture and environmental management, particularly in large river basins, it is important to model simultaneous sequences of multiple variables (e.g., precipitation and temperature) over large heterogeneous areas, while maintaining physically plausible spatial, temporal and inter-variable relationships. Several approaches have been developed for simultaneous multisite multivariate generation of climate variables but those are still inadequate to model the joint distribution of, for example, precipitation and temperature simultaneously at all sites as well as inter-variable and inter-site dependence structures.

In this chapter, the suitability of GLMs for multisite multivariate modelling of precipitation and temperature fields in the Canadian Prairie Provinces, with the aim of using these models for downscaling GCM outputs for climate change impact analysis, is assessed. This chapter contains the following submitted manuscript:

1. **Asong ZE**, Khaliq MN, Wheater HS (2015), Multisite multivariate modelling of daily precipitation and temperature in the Canadian Prairie Provinces using Generalized Linear Models. Submitted to *Journal of Climate Dynamics*, paper # CLDY-D-14-00659 (status: "under review" as of 14/12/2015).

Abstract

Based on the Generalized Linear Model (GLM) framework, a multisite stochastic modelling approach was developed using daily observations of precipitation and minimum and maximum temperatures from 120 sites located across the Canadian Prairie Provinces: Alberta, Saskatchewan and Manitoba. Temperature was modeled using a two-stage normal-heteroscedastic model by fitting mean and variance components separately. Likewise, precipitation occurrence and conditional precipitation intensity processes were modeled separately. The relationship between precipitation and temperature was accounted for by using transformations of precipitation as covariates to predict temperature fields. Large scale atmospheric covariates from the National Center for Environmental Prediction Reanalysis-I, teleconnection indices, geographical site attributes, and observed precipitation and temperature records were used to calibrate these models for the 1971–2000 period. Validation of the developed models was performed on both pre- and post-calibration period data. Results of the study indicated that the developed models were able to capture spatiotemporal characteristics of observed precipitation and temperature fields, such as inter-site and inter-variable correlation structure, and systematic regional variations present in observed sequences. A number of simulated weather statistics ranging from seasonal means to characteristics of temperature and precipitation extremes and some of the commonly used climate indices were also found to be in close agreement with those derived from observed data. This GLM-based modelling approach will be developed further for multisite statistical downscaling of Global Climate Model outputs to explore climate variability and change in this region of Canada.

Keywords: GLMs; extreme events; precipitation; stochastic modelling; temperature; weather generators

3.1 Introduction

Assessment of vulnerability of local and regional water supply protection and management projects to climate change is currently an active area of research. In this context, many impact related studies tend to rely on outputs from Global Climate Models (GCMs). However, these outputs cannot be applied directly at local and regional scales primarily due to the spatial resolution of GCMs, which is much coarser than that typically required for many impact assessment studies (Fowler et al., 2007; Maraun et al., 2010). Also, some investigators have expressed doubts about the reliability and local scale utility of some GCM outputs (e.g., precipitation) that are critically dependent on sub-grid scale processes such as those involving clouds and topography (Huth, 2002; Cavazos and Hewitson, 2005; Dibike et al., 2008). These limitations lead to a scale mismatch between the information that GCMs at the moment are able to provide and that which is desired in many impact assessment studies (e.g., Zorita and von Storch, 1997).

To circumvent the above mentioned shortcomings, techniques based on dynamical and statistical downscaling have emerged. The dynamic downscaling techniques use Regional Climate Models (RCMs) to predict finer-scale climate variables when these models are driven by GCM outputs at their boundaries (Giorgi, 2006). Though on the rise, direct application of RCM outputs for regional impact assessment is often restricted because of the high computational cost involved and/or bias partly originating from the driving GCM. Alternatively, statistical downscaling aims at relating large scale atmospheric covariates to local scale surface variables (Wilby and Wigley, 1997; Turco et al., 2011; Gutiérrez et al., 2013; D'Onofrio et al., 2014). One of such techniques is the use of weather generators (WGs) for simulating long sequences of weather variables. The simulated sequences capture essential features of the observations. A

detailed review of these techniques, which generally fall under the category of stochastic modelling tools, can be seen in Maraun et al. (2010).

Stochastic modelling of weather variables at the daily or sub-daily scale is particularly challenging due to the intermittence that is inherent in, for example, precipitation at such scales. In some studies, precipitation has been modeled by a two-stage process involving separate models for precipitation occurrence and amounts when wet (Todorovic and Woolhiser ,1975; Katz, 1977; Buishand, 1978; Stern and Coe, 1984). Daily precipitation occurrence is often modeled using a two-state Markov process corresponding to wet and dry states (e.g., Richardson, 1981; Wilks, 1998; Katz et al., 2003), while the gamma distribution has commonly been used to model precipitation amounts (Katz, 1977; Stern and Coe, 1984). Elsewhere, exponential and mixed exponential distributions (Richardson 1981), as well as mixtures of different continuous distributions (Hundecha et al., 2009) have been used. Generalized Linear Models (GLMs) (McCullagh and Nelder, 1989) offer a framework that unites and extends many of the existing approaches that have been proposed to model precipitation occurrences. These models have been utilized successfully for modelling precipitation sequences (e.g., Chandler and Wheater, 2002; Furrer and Katz, 2007).

Chandler and Wheater (2002) used a logistic regression to model the probability of rain on a given day at stations in Ireland, with the observed North Atlantic Oscillation being the predictor explicitly representing the large scale atmospheric structure in addition to an indicator of precipitation occurrence on the previous day. They found such models to provide a good representation of the organized structures in the precipitation data in addition to satisfying their distributional assumptions. Kenabatho et al. (2012) explored GLMs to model daily rainfall data from 13 stations located in the Limpopo basin in Botswana. Although their results showed quite high uncertainty, they recommended GLMs for modelling rainfall sequences in semi-arid

climates. In the Peruvian Andes, Bergin et al. (2012) modeled daily rainfall using GLMs and concluded that rainfall statistics were satisfactorily reproduced by the models particularly in relatively small catchments. In the context of multisite daily rainfall downscaling in Australia (Frost et al., 2011), the performance of the GLM-based WG was found quite satisfactory compared to other state-of-the-art techniques. Recently, Chun et al. (2013) performed a comparative single-site downscaling of daily precipitation at four selected locations in western Canada using the LARS-WG (Long Ashton Research Station weather generator) (Semenov and Stratonovitch, 2010) and GLM approaches. Although both approaches were able to reproduce most of the statistical properties of the historical precipitation records, the GLM-based WG out-performed the LARS-WG in terms of simulating characteristics of extreme events as well as inter-annual variability of precipitation sequences.

Most of the WGs focus on individual sites (e.g., Rajagopalan and Lall, 1999; Wilby et al., 2002) and are therefore unable to represent the spatial structure of the observed climatic variables. Although these models can generate time series at more than one site when applied separately, the series so generated would not be spatially consistent, due to neglecting inter-station correlations (Wilks, 1998; Mehrotra and Sharma, 2007; Jeong et al., 2012). However, for many water resources design and management related projects, particularly in large river basins, it is important to model simultaneous sequences of multiple variables (e.g., precipitation and temperature) over large heterogeneous areas, while maintaining physically plausible spatial, temporal and inter-variable relationships. Several approaches have been developed for simultaneous multisite multivariate generation of climate variables (Apipattanavis et al., 2007; Steinschneider and Brown, 2013). However, as noted by Maraun et al. (2010), multisite generation offers many significant challenges primarily due to the need to model the joint distribution of, for example, precipitation simultaneously at all sites and inter-variable and inter-

site dependence structures. The GLM-based Rglimclim software package (Chandler, 2014) provides a flexible framework for accomplishing such tasks within the R programming environment (R Development Core Team, 2014).

This study seeked to investigate the suitability of GLMs for multisite multivariate modelling of precipitation and temperature fields in the Canadian Prairie Provinces, with the aim of using these models for downscaling GCM outputs for climate change impact analysis. This region comprises 47 diverse watersheds including the Saskatchewan, Athabasca, Peace and Churchill River Basins, which serve various needs of the communities ranging from agricultural to domestic usage and fulfilling rapidly expanding requirements of the industrial sector. Apart from regional inhomogeneity and a paucity of ground-based observations, this region of Canada is also characterized by a highly variable hydro-climate with recurrent floods and multi-year droughts.

The paper is organized as follows: Section 3.2 describes the study area and datasets used. The methodology for multisite multivariate modelling of precipitation and temperature sequences based on the *Rglimclim* software package is described in Section 3.3. Results of the study are presented and discussed in Section 3.4, while a summary and conclusions are given in Section 3.5.

3.2 Datasets Used

The datasets used in this study included daily total precipitation, and minimum and maximum temperatures for the 1961 to 2005 period from a network of 120 stations (Figure 3-1, and Table A-1 in Appendix A), obtained from Environment Canada (http://www.ec.gc.ca). Temperature is recorded at 96 of the 120 stations. These datasets have been quality controlled and adjusted to account for known changes in recording practice (see Vincent, 2009; Mekis and Vincent, 2011).

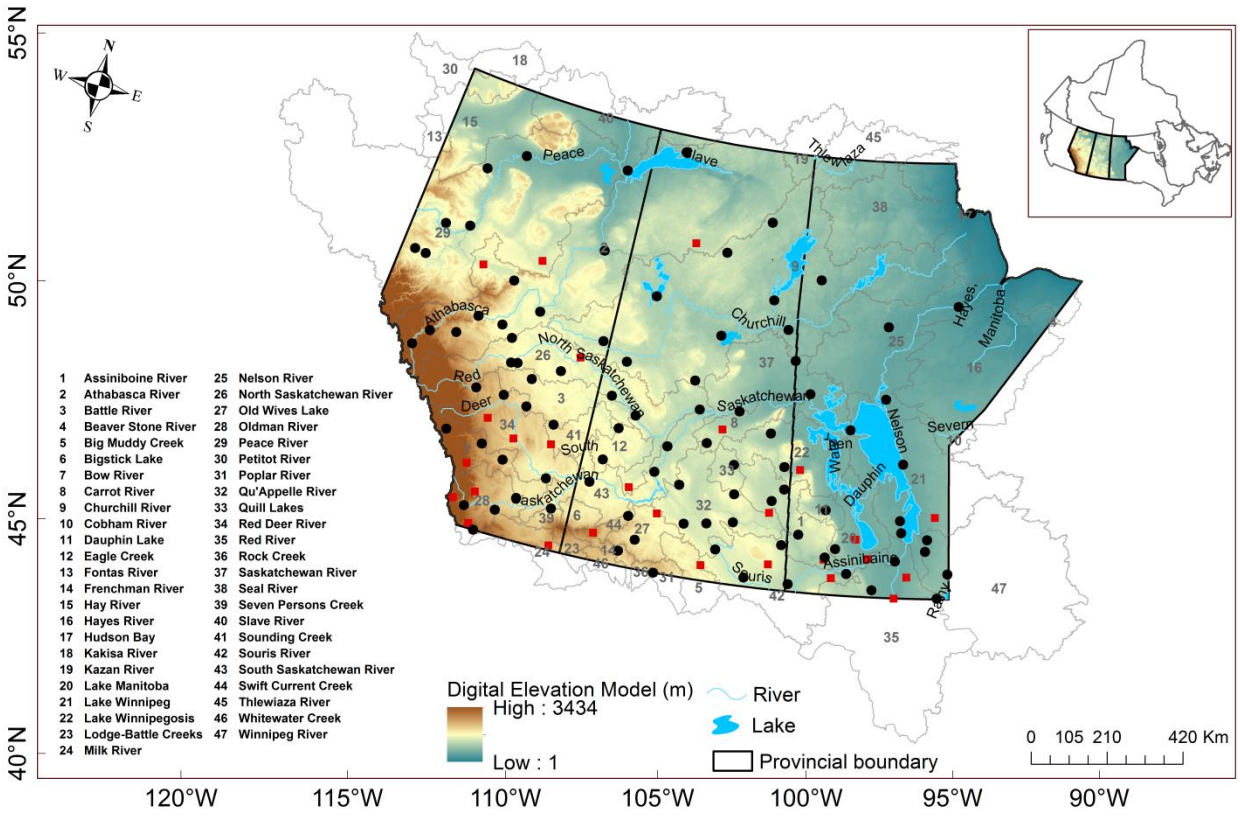

Figure 3-1: Study area and observation stations (black dots and red squares) considered in the study. Precipitation is observed at all stations, while temperature is recorded only at stations indicated as black dots. Forty seven watersheds spanning the study area including the provinces of Alberta, Saskatchewan and Manitoba (left to right) are also shown. The inset shows location of the study area in Canada.

Standardized daily values of large scale atmospheric covariates were derived for the 1961–2005 period from the National Center for Environmental Prediction and the National Center for Atmospheric Research (NCEP/NCAR) Reanalysis-I (Kalnay et al. 1996) over a spatial domain encompassing latitudes 40^{o}N to 70^{o}N and longitudes 130^{o}W to 70^{o}W. In total, 21 large scale covariates (wind speed at 10-m, 500- and 850-hPa; U-component and V-component at 10-m, 500- and 850-hPa, vertical velocity, geo-potential height, specific humidity, and relative humidity at 850- and 500-hPa; total cloud cover, mean sea level pressure, precipitable water and 2-m air temperature) were explored. Monthly indices of teleconnection patterns, such as Pacific Decadal Oscillation (PDO), Pacific North American mode (PNA) and Artic Oscillation (AO),

were sourced from the Joint Institute for the Study of the Atmosphere and Ocean, University of Washington (http://jisao.washington.edu/analyses0302/).

It is important to note that the above mentioned observed temperature and precipitation datasets, large scale atmospheric covariates and indices of PDO, PNA and AO were used in Asong et al. (2015) to partition the study area into five homogeneous precipitation regions on which most of the analyses presented herein were based. The partitioning was done using the same set of atmospheric covariates as are used in the present study.

3.3 Methodology

This section provides methodological background of the GLM framework for modelling daily precipitation and temperature variables. In addition, other important topics ranging from selection of covariates, spatial-temporal dependence structure to model calibration and validation procedures are also discussed. The methodology is described as implemented in the Rglimclim software package of Chandler (2014), which was used for this study.

3.3.1 GLM for Daily Precipitation

A two-stage approach involving separate amount and occurrence models has been used previously to model precipitation sequences (Coe and Stern 1982; Chandler and Wheater, 2002; Chandler, 2005; Furrer and Katz, 2007). In a GLM, an $n \times 1$ vector of data $y_1, ..., y_n$ are considered to be the realized values of the random variables $\boldsymbol{Y} = (\boldsymbol{Y_1}, ..., \boldsymbol{Y_n})'$ with a mean vector $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)'$ where μ_i is related to the values of a row vector $\boldsymbol{x_i}$ of predictors such that:

$$g(\mu_i) = \boldsymbol{x_i}\boldsymbol{\beta} = \eta_i \tag{3-1}$$

where $g(.)$ is a monotonic transformation known as the link function and $\boldsymbol{\beta}$ is a $x \times 1$ vector of coefficients. The precipitation occurrence process (i.e., the pattern of wet and dry days) is modelled using logistic regression and the precipitation amounts (i.e., intensity) process on wet

days is modelled using the gamma distribution. The precipitation occurrence process takes the form:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{x_i}\boldsymbol{\beta} \tag{3-2}$$

where p_i is the probability of precipitation for the i^{th} case in the dataset conditional on a covariate row vector $\boldsymbol{x_i}$ with coefficient column vector $\boldsymbol{\beta}$. Subsequently, for a potentially different covariate vector $\boldsymbol{\xi_i}$, the precipitation intensity process for the i^{th} wet day is modelled as gamma-distributed with mean μ_i and shape parameter ν, where

$$\ln(\mu_i) = \boldsymbol{\xi_i}\boldsymbol{\varphi} \tag{3-3}$$

with the shape parameter ν assumed to be constant (e.g., Yang et al., 2005) for all observations at all sites, and $\boldsymbol{\varphi}$ is a column vector of coefficients. The coefficient vectors $\boldsymbol{\beta}$ and $\boldsymbol{\varphi}$ are estimated using the maximum likelihood method assuming that the observations from different sites are independent (Chandler, 2005; Chandler and Bate, 2007), with subsequent adjustments for inter-site dependence that is generally present.

3.3.2 GLM for Daily Temperature

Khalili et al. (2013) developed a statistical downscaling approach to model daily minimum (*Tmin*) and maximum (*Tmax*) temperatures at 10 different locations in Ontario and Quebec. Their approach consists of a combination of a linear regression component to describe the linkage between predictors and temperature values, and a stochastic component based on a spatial moving-average process to reproduce the observed spatial dependence between the values at different sites. Several other approaches also exist in the literature. For example, regression-based methods and artificial neural networks were used by Schoof and Pryor (2001), while first–order trivariate auto-regression that is conditional on precipitation occurrence as implemented in Weather GENerator (WGEN) by Richardson and Wright (1984) have also been applied

extensively. Elsewhere, Chen et al. (2012) developed the MulGETS WG wherein a first–order auto-regression was used to model temperature, while Furrer and Katz (2007) modelled both precipitation and temperature at multiple sites using GLMs. Standard linear regression methods assume constant variance for daily time series, $\boldsymbol{Y}_{st}$, at each site s on a given day t. However, the assumption of constant variance is often violated when analyzing temperature series at the daily time scale (Chandler, 2014). Therefore, following Chandler (2005), the method used here included a two-stage approach whereby separate mean and variance components were developed within a normal-heteroscedastic framework in which the mean (μ_{st}) and variance (σ^2) of $\boldsymbol{Y}_{st}$ depended on possibly different covariate vectors. For modelling *Tmin* and *Tmax*, we began by modelling the mean of the two variables using a normal distribution, and then the difference between them using a gamma distribution. This guaranteed that *Tmax* is always greater than *Tmin* in simulated sequences.

3.3.3 Selection of Probable Candidate Predictors

Selection of significant candidate predictors is the most important factor that could affect the accuracy of the estimated predictands (Wilby and Wigley, 2000). Recently, Asong et al. (2015) studied spatio-temporal relationships of various precipitation characteristics and the predictors described above in Section 3.2. Principal component and canonical correlation analyses were used to screen the large scale covariates. They found the following eight predictors to influence significantly the precipitation characteristics both in space and time: 2-m air temperature, 850-hPa relative humidity, 500-hPa specific humidity, 850-hPa geo-potential height, mean sea level pressure, horizontal wind components (850-hPa meridional and 10-m zonal wind), vertical velocity (i.e., omega at 500-hPa), and the PDO and PNA indices. The selected predictors reflected information about the thickness, circulation and moisture content of

the atmosphere. Subsequently, for modelling precipitation, *Tmax* and *Tmin*, the statistical significance of the covariates was assessed simultaneously using likelihood ratio tests, adjusted for inter-site dependence following the approach described in Chandler and Bate (2007), when extending a model by adding more covariate terms in the GLM framework. Thus, ensuring parsimony and reducing the artefacts resulting from over-fitting.

3.3.4 Spatial-Temporal Dependence Structure and Distributional Assumptions

Daily weather sequences often exhibit a high level of temporal and spatial autocorrelation (Wilks, 1998). The GLM framework allows for modelling of marginal distributions. However, the flexible approach of *Rglimclim* offers an opportunity for incorporation of several inter-site dependence models. Given that most weather sequences at different sites tend to be correlated, potentially as a result of being produced by similar large scale weather systems, it is possible to construct a joint distribution of precipitation or temperature at all sites which respects marginal distributions from at-site GLMs. A meaningful GLM for generating multisite multivariate weather sequences must therefore preserve the spatial coherence. This requires a computationally tractable representation of inter-site dependence. This feature was incorporated by transforming the precipitation amounts to Gaussianity and then studying inter-site correlations on the transformed scale (see Yang et al., 2005 for details). For temperature, inter-site dependence was specified directly via correlations between the standardized residuals. The software also offers various options for modelling temporal autocorrelation structure mostly as a function of lagged values and a 'persistence indicator'. Intervariable relationships were represented as functions of concurrent/simultaneous and lagged values of other variables. An advantage of using a spatial correlation model is that it provides the opportunity to simulate weather sequences at ungauged locations which is an important consideration for the current study area due to the sparse network of observation stations.

Multisite simulation of precipitation occurrence in a large study area with marked convective activity during summer makes the incorporation of spatial dependence into binary sequences a very challenging task. Yang et al. (2005) reviewed related techniques in the context of daily rainfall generation and found that none of the approaches was suitable for their study case. Their main difficulty was that the study area was relatively small compared to the synoptic weather systems affecting it. As our study area is very large and the precipitation production processes (e.g., convective cells) are highly localized, we adopted the same approach as in Ambrosino et al. (2014). Supposing that it is required to generate a vector $\boldsymbol{Y} = (Y_1, ..., Y_{st})'$ of correlated binary variables and that Equation 3-2 gives the probability of precipitation at site st as p_{st} . A conceptually easy to implement approach is to start by generating a set of correlated Gaussian variables $\boldsymbol{Z} = (Z_1, ..., Z_{st})$ and then define a threshold that is chosen to ensure that $P(Y_{st} = 1) = p_{st}$ as required by the logistic regression model in Equation 3-2.

3.3.4 Model Fitting and Evaluation: Calibration and Validation

The primary stage in model building was to decide on an appropriate class of models to represent the variable(s) of interest, which was addressed in sub-sections 3.3.1 and 3.3.2 above in the context of GLM framework. In this study, GLMs were fitted separately to precipitation and temperature fields (i.e. *Tmin* and *Tmax*) considering the entire study area as a single region and using observations from the 1971–2000 period. Herein, a day was defined as wet if the recorded amount of precipitation exceeded 0.5 mm. First, for the precipitation case, models were fitted using data from all 120 sites. Subsequently, *Tmin* and *Tmax* from 96 of the 120 stations were modeled separately and intervariable relationships were accounted for by using simultaneous and lagged values of precipitation as covariates to model temperature. This approach was refined further based on smaller homogeneous partitions of the study domain. The first step involved in

the calibration was the development of 'initial' GLMs consisting of a constant term and basic factors influencing weather variability such as seasonality, autocorrelation and geographical attributes (site effects). Subsequently, daily values of NCEP-based covariates and monthly values of teleconnection indices (see sub-section 3.3.3) were incorporated as external covariates. The rationale for adding successive predictors to the existing model was assessed by evaluating the predictive performance, dependence-adjusted log-likelihood and the residual structure for each fitted model.

It is possible, for example, that climate variability in the Canadian Prairie Provinces is linked with the PDO and PNA phenomena, especially during winter months. Therefore, the coefficient of the PDO in a GLM should vary by season of the year. Instead of fitting separate models for each month of the year, the coefficient of the PDO can be represented as a linear combination of covariates explaining seasonality. This was achieved within the GLM framework via interactions (Chandler and Wheater, 2002; Chandler and Scott, 2011). The software provides a wide range of residual-based diagnostics to check that the fitted models are able to reproduce the systematic structure in the observations, as well as the distributional assumptions (e.g., precipitation intensities follow gamma distributions) and the assumed inter-site correlation structure (see Yang et al., 2005 for further details). For example, to check that the underlying structure has been captured by the fitted model, we define Pearson residuals as:

$$r_i^{(P)} = \frac{Y_i - \mu_i}{\sigma_i} \tag{3-4}$$

Where Y_i is the observed response for case i, and μ_i and σ_i are the modeled mean and standard deviation. If the fitted model is correct, all of the Pearson residuals have expected value zero and variance 1. In addition to Pearson residuals, Anscombe residuals (Equation 3-5) for the gamma

distribution are defined for the amounts model to ensure that the probability structure of the fitted models is correct.

$$r_i^{(A)} = \left(\frac{Y_i}{\mu_i} \right)^{1/3} \tag{3-5}$$

The suitability of the calibrated models for generating weather sequences independent of the calibration period was tested by validating the models on the pre- and post-calibration periods (i.e., 1961–1970 and 2001–2005). To simulate weather sequences, the parameters of the fitted models were constrained using external covariates from the corresponding validation periods. For comparing simulated statistics with observed ones, it is important to assess the uncertainty resulting from missing observations. For this purpose, 39 imputations (whereby missing values at gauged and ungauged sites are sampled from their conditional distributions given the available observed data; see Chandler, 2014, page 64 for details) for defining the 95% uncertainty interval for the true value were carried out using predictors from the respective calibration and validation periods. Selected statistics, such as the *Mean*, standard deviation (*Std*), lag-1 autocorrelation function (*ACF*(1)), proportion of wet days (P_W), conditional mean ($Mean^{cond}$) and conditional standard deviation (Std^{cond}) were computed for each of the resulting imputed data sets. The variability in the resulting statistics is indicative of the historical uncertainty due to missing values. Conditional statistics were computed for precipitation only, based on the proportion of exceedances of the 0.5 mm threshold. Using the fitted models, 100 realizations were obtained for the calibration and validation periods. In each case, predictors for the first year were used to initialize simulations. Subsequently, the same selected statistics were computed from the simulated sequences and compared with the corresponding observed values. Model performance was first evaluated by region and then by site.

3.3.4.1 Additional assessments

It is likely that changes in the seasonal and extreme precipitation characteristics will have important implications for managing regional water resources related projects in the study area (Mladjic et al., 2011; Khaliq et al., 2014). Therefore, in addition to the above mentioned statistics, the ability of the GLMs in reproducing observed distributions of seasonal extremes was also assessed. For this purpose, seasonal maxima (minima) of daily *Tmax* (*Tmin*) were derived from observed data as well as from simulated sequences for the calibration and the two validation periods. In like manner, seasonal maxima of daily precipitation amounts were obtained from the observed and simulated data. For example, for the 120 sites for the calibration period (1971–2000), 100 simulations of precipitation per site were made, and then for each season, the maximum value per year was extracted for each simulation and for a given site. This gave 30 maxima/minima per year per simulation. Then, the 95th percentile value was computed from each simulation, resulting to one value per simulation. Subsequently, the 95th (5th) percentile of observed precipitation and *Tmax* (*Tmin*) extremes was compared to the 100 95th percentiles values obtained from 100 simulations. In addition, seasonal values of commonly used climate indices, i.e., mean wet spell length (pwsav), mean dry spell length (pdsav), maximum number of consecutive dry days (pxcdd), maximum number of consecutive wet days (pxcwd), and extreme hot and cold temperature spells (i.e. the 90th percentile heat wave duration–txhw90 and the 10th percentile cold wave duration–tncw10), were investigated. These indices have been selected from a set of 27 different indices suggested by Goodess (2003) in order to develop a set of harmonized indices across the globe. Specifically, for txhw90, let Tx_{ij} be the daily maximum temperature at day *i* of period *j* and let $Txq90_{inorm}$ be the calendar day 90th percentile calculated for a 5-day window centered on each calendar day during a specified period. Then the maximum

number of consecutive days per period, where $Tx_{ij} > Txq90_{inorm}$, was obtained to calculate txhw90. Similarly, for tncw10, let Tn_{ij} be the daily minimum temperature at day *i* of period *j* and let $Tnq10_{inorm}$ be the calendar day 10th percentile calculated for a 5-day window centered on each calendar day during a specified period. Then the maximum number of consecutive days per period, where $Tx_{ij} < Txq90_{inorm}$, was obtained to calculate tncw10. Further details on the computation of other indices can be found at http://www.cru.uea.ac.uk/projects/stardex/deis/Diagnostic_tool.pdf.

3.4 Results and Discussion

This section contains results of various components of the study, ranging from preliminary diagnostics to model calibration and validation, as well as associated discussions. Though all components are presented and discussed in separate sections, graphical outputs of the validation part of the study are presented alongside the calibration results for ease of comparison.

3.4.1 Preliminary Diagnostics, Inferences, and Calibration of GLMs

We started by fitting GLMs to precipitation sequences from all 120 sites, by considering the entire study domain as a single region, and then diagnose Pearson residuals, classified by site, month and year, for the presence or absence of unexplained spatiotemporal structures. Following this approach, the spatial distributions of "mean residuals by site" obtained from the amounts and occurrence models for all sites are shown in Figure 3-2. In the presence of any systematic regional variations that are not accounted for by the fitted model, the sites with positive mean residuals will tend to cluster together and the same will be the case for negative mean residuals. In Figure 3-2, a discernible spatial trend in the pattern of residuals is evident. For example, to the southeast and in western parts of the study domain, clusters of positive-only residuals (unfilled circles) can be seen. Likewise, to the south-central region, groupings of negative-only residuals

are evident. Additional results of the residual analysis by month and year for the same amounts and occurrence models (see Figure B9) suggested that a single model for the entire region was not adequate for describing daily precipitation sequences because the pattern of residuals do not satisfy the underlying distributional assumptions. Moreover, it was also noted that most of the selected statistics and inter-variable correlations were not satisfactorily reproduced.

Having gained insights from the results discussed above, GLMs were fitted separately to each of the five pre-defined statistical and climatological homogeneous partitions/regions of the study area, identified recently in Asong et al. (2015) (Figure 3-3). These regions were delineated using principal component and canonical correlation analyses and Fuzzy C-Means clustering of the feature vectors derived from large scale atmospheric covariates and geophysical attributes. The pattern of residuals shown in Figure 3-2 shows some similarity with the geographical extent of these homogeneous regions. Therefore, the rest of the analyses for the precipitation case presented hereafter are based on models fitted separately to each of these regions. Evaluation of the residuals from the fitted models for each region indicated a good fit, when assessed on the basis of 95% confidence intervals (see Appendix B).

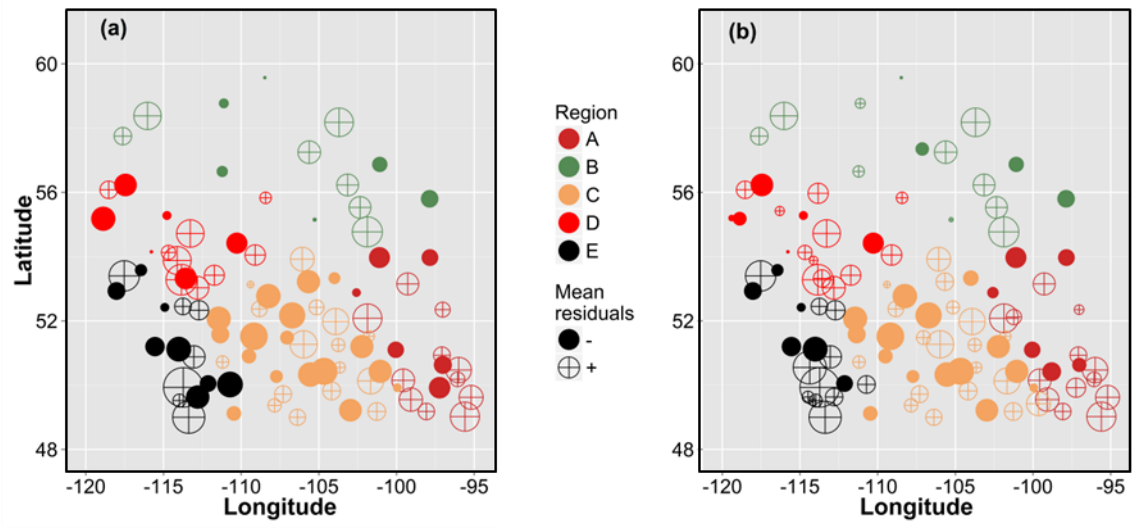

Figure 3-2: Bubble map showing spatial distribution of mean Pearson residuals at each site from the fitted precipitation (a) amounts and (b) occurrence models. The bubble maps were obtained from the GLMs fitted by considering the entire study domain as a single region. The size of the circle is proportional to the standardized mean residual. Description of the regions A to E corresponding to different colors is provided in Figure 3-3

For the temperature field, *Tmin* and *Tmax* were modeled separately considering the entire study domain as one region. Based on the residual plots, distributional features of both *Tmin* and *Tmax* are relatively better described by the GLMs compared to the precipitation field when the entire domain was considered as one region. To develop a joint model for precipitation and temperature, we used concurrent and/or lagged precipitation values in each homogenous region as a covariate to model temperature.

The influence of teleconnections on regional precipitation and temperature patterns was also examined. The PDO and PNA were found to be the dominating teleconnection indices modulating regional and seasonal precipitation patterns. Spatially, the PDO was found to influence significantly precipitation processes in the western and northeastern parts of the study area, while the PNA showed dominance in the southeast (region A in Figure 3-3). Temporally, the PDO and PNA were found to have a substantial time-lag for precipitation occurrence and intensity processes for up to three years for most parts of the study area. However, a simultaneous response was found between the PDO and variance of *Tmin* and *Tmax*. Given that no simultaneous response was found between precipitation and teleconnection indices, it is likely that the atmospheric patterns delivering precipitation over the study region are not closely associated with the atmospheric patterns that control PDO and PNA variations.

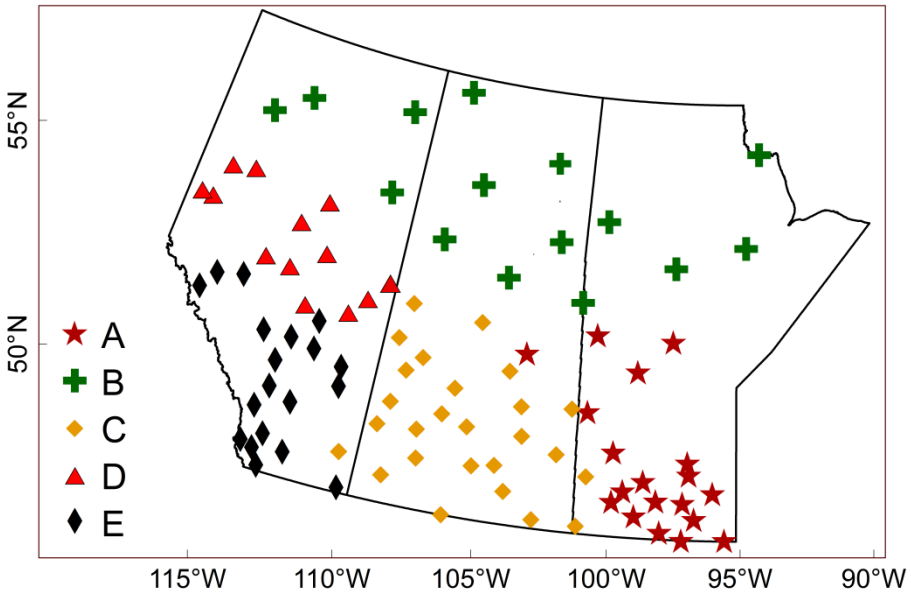

Figure 3-3: Statistical and climatological homogeneous regions (A, B, C, D and E), along with the spatial distribution of respective defuzzified precipitation gauges from Asong et al. (2015)

3.4.1.1 Evaluation of spatial dependence and distributional assumptions

The ability of the GLMs to preserve the probability structure of the observed precipitation and temperature fields was assessed through Q–Q plots of standardized Anscombe's residuals under the fitted amounts models. Besides, the relationship between the observed and modeled inter-site correlations with distance, calculated from the site's latitude and longitude was also examined. A powered exponential correlation function with decreasing correlation at large distances (Chandler, 2014) was found suitable for modelling inter-site dependence of conditional precipitation intensity process, and temperature values. Figure 3-4 shows the fitted correlation models for each region, alongside Q–Q plots of the residuals pooled over all sites in each region. For all regions, the residuals correspond to the theoretical values very well (Figure 3-4a). Figure 3-4b shows observed inter-site correlations, overlain by the fitted models. The exponentially decaying behavior of observed correlations is well described by the assumed theoretical models. In summary, inter-site correlations for all regions are well captured. In Figure 3-5, results of spatial dependence analysis for temperature are shown. The upper row corresponds to *Tmin*

while the lower one shows plots for *Tmax*. The fitted inter-site correlations generally are in good agreement with those observed and are judged to be satisfactory for additional analyses. However, slight discrepancies for the lower end of the distribution can clearly be noted.

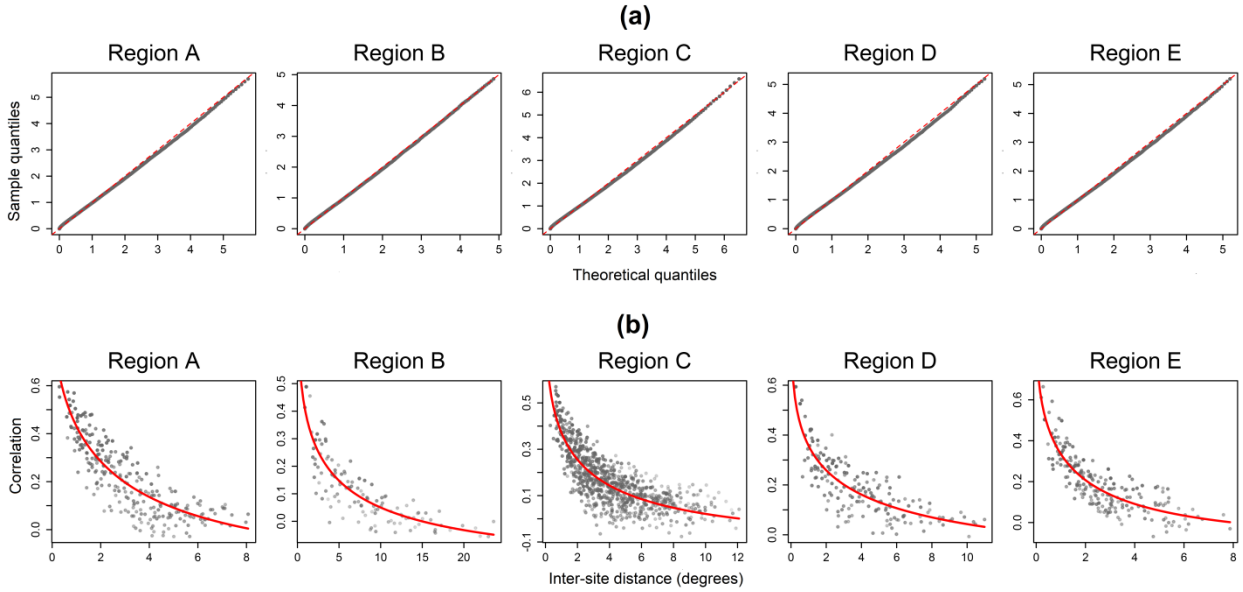

Figure 3-4: (a) Q–Q plots of standardized Anscombe residuals pooled over all sites in each region, for the fitted precipitation amounts model; (b) Observed inter-site correlations and the fitted correlation model (red line)

3.4.1.2 Simulated characteristics of daily, seasonal and extreme values of precipitation and temperature

Figures 3-6 and 3-7 show regionally pooled (i.e., over all sites in a region) simulated values of selected statistics (see sub-section 3.3.4) of daily precipitation, together with simulated distributions obtained from 39 imputations in lieu of missing observations. In Figure 3-6, there is generally a good agreement between the simulated and observed values for each month of the year, with few exceptions. The observed values (i.e., grey) of P_W are slightly overestimated for nearly all regions, particularly for the summer months. The performance of the models for P_W appears to be sensitive to the choice of the threshold used for defining a wet day because the values of P_W were found to be relatively well reproduced when 1 mm (instead of 0.5 mm) threshold was used. Though with a wider simulated distribution, the $ACF(1)$ values were also

satisfactorily reproduced for all regions and months, except for the month of December and about same is the behavior of the $Mean^{cond}$. Based on the analyses performed for other sites (See for example Figure 4-3, 4-4 and 4-5), the GLMs were able to reproduce the systematic regional variations and spatial structures of both mean and extreme weather states at the majority of the 120 sites.

For some applications such as water balance studies, it is important to reproduce observed variations in precipitation totals over monthly or longer time scales. Moreover, simulating the inter-annual variability is particularly important as it indicates that the model is correctly reproducing the predictor-predictand relationships. This feature provides some additional confidence that changes in the predictors under climate-change conditions will be able to produce correct changes in the predictands (Haylock et al., 2006) when these models will be used in that context. Figure 3-7 shows simulated values of seasonal mean daily precipitation and the corresponding observed values for three selected regions A, B and C. Apart from region A and C, where there is a slight tendency for the model to overestimate the monthly precipitation totals in spring and summer, the GLM framework is able to preserve the observed inter-annual variability. This feature of the GLM framework is also discussed in Chun et al. (2013). For most of the years, the observed precipitation values are found to be within the 2.5^{th} and 97.5^{th} percentiles of the simulated distribution. Also for the case of observed precipitation, the imputation range (i.e., the thick grey band) points to substantial uncertainty due to missing values; for example, region B in winter. The behavior of the seasonal mean precipitation for the remaining two regions D and E was about the same as discussed above for region B.

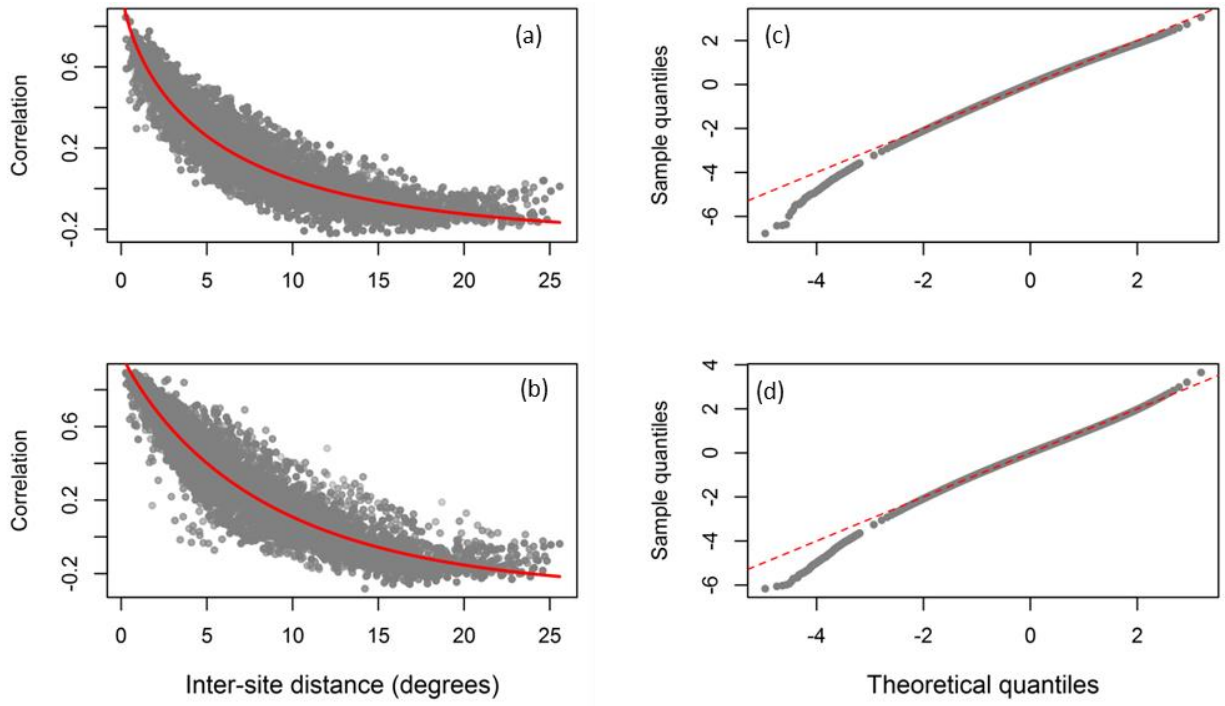

Figure 3-5: Inter-site correlations (grey dots) that decay exponentially with distance for daily (a) minimum and (b) maximum temperatures. Q–Q plots of standardized Anscombe's residuals from the jointly fitted mean and variance model for daily (c) minimum and (d) maximum temperatures, respectively

The next assessment was for the probability distribution of monthly precipitation amounts. One way to assess the ability of a model in simulating the probability distribution of observed monthly precipitation totals is by plotting quantiles of simulated and observed amounts. Figure 3-8(a) shows Q–Q plots of observed and simulated monthly precipitation totals averaged over the number of sites in each region. These plots indicate a good correspondence between observed and simulated monthly precipitation totals for all regions for the calibration period.

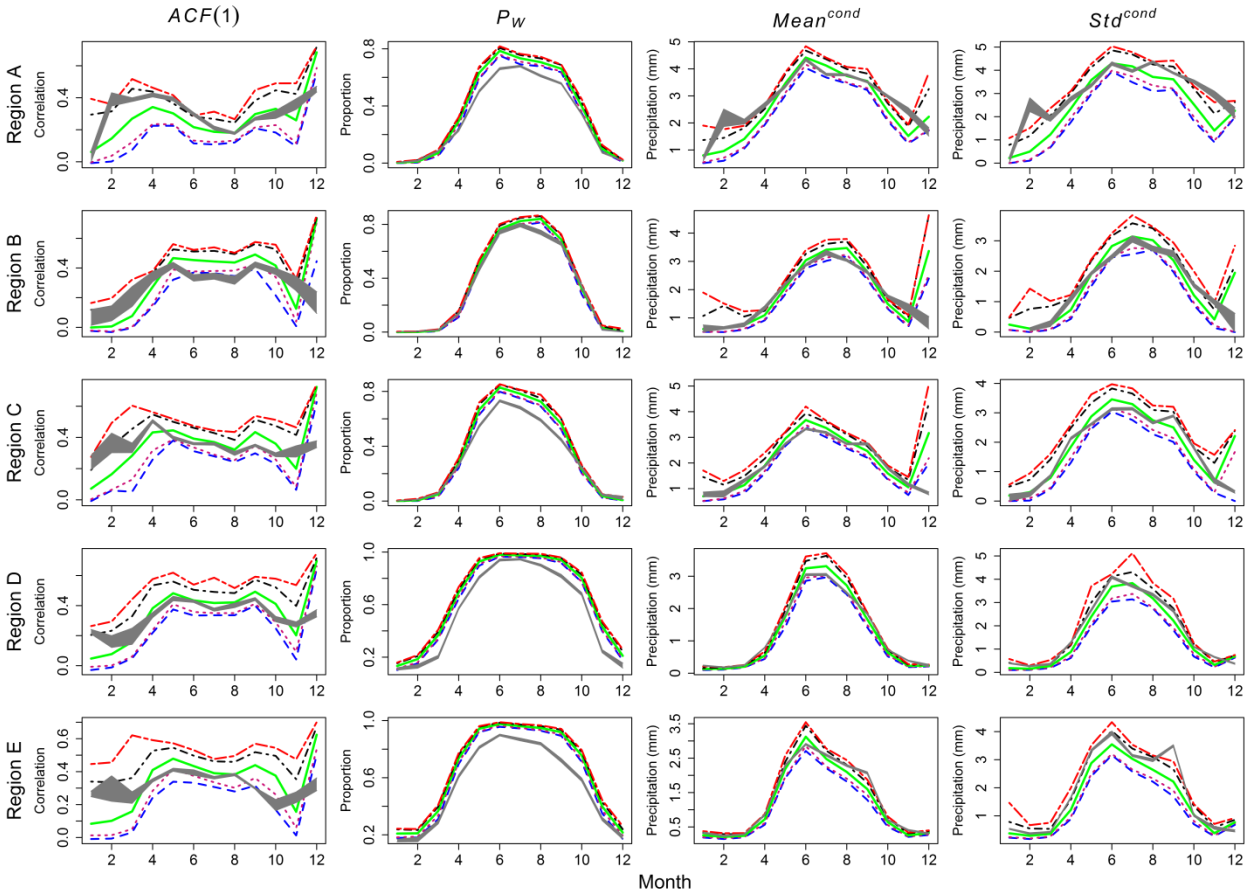

Figure 3-6: Comparison of observed and simulated values of selected statistics – lag-1 autocorrelation function (*ACF*(1)), proportion of wet days (P_W), conditional mean ($Mean^{cond}$), and conditional standard deviation (Std^{cond}) of precipitation sequences – for all regions for the calibration period (1971–2000), together with distributions obtained from 39 imputations of observed data. Thick grey band is the 95% interval for the imputed values. The pink, green and black lines indicate respectively the 2.5th, 50th and 97.5th percentiles, while the blue line represents the minimum and the red line represents the maximum values of the simulated precipitation amounts

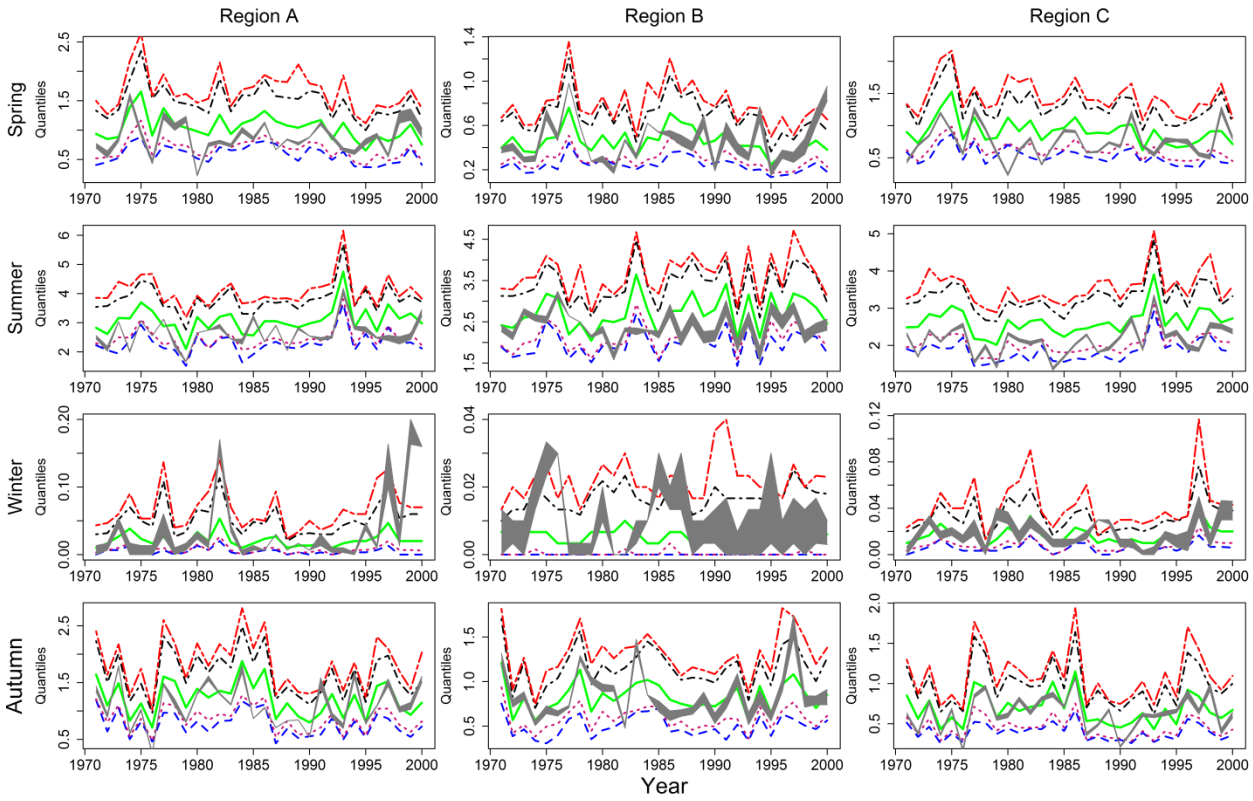

Figure 3-7: Comparison of observed and simulated values of spring (MAM), summer (JJA), winter (DJF) and autumn (SON) daily precipitation pooled over all sites in a region. Results are shown for three selected regions A, B and C and remaining convention is the same as in Figure 3-6

Figures 3-9(a) and 3-10(a) show monthly statistics of *Tmax* and *Tmin*, respectively. Unlike precipitation, most statistical properties of both temperature fields were well reproduced by the GLMs, except *ACF*(1) which was underestimated for summer months. The last two columns in these Figures show intervariable correlations. For *Tmax*, the correlation between its selected percentiles and that of the precipitation field were not well reproduced especially in summer. This could be due to the inability of the GLMs to capture the localized, short duration and generally more intense convective precipitation during the summer period. This may also be due to the non-linear coupling between point-scale precipitation and large scale atmospheric covariates which may not have been captured by the regression-based linear approach used here. To overcome this shortcoming, the use of Convective Available Potential Energy (CAPE) and

Convective Inhibition (CIN) which play a dominant role in convective precipitation (both its genesis and intensity) as additional external covariates in the GLMs can be explored in future studies. Unlike inter-variable correlations of precipitation and temperature fields, the correlation between *Tmax* and *Tmin* was fairly well captured for most months, except the winter months. Compared to the correlation between precipitation and *Tmax*, the correlation between precipitation and *Tmin* was satisfactorily reproduced.

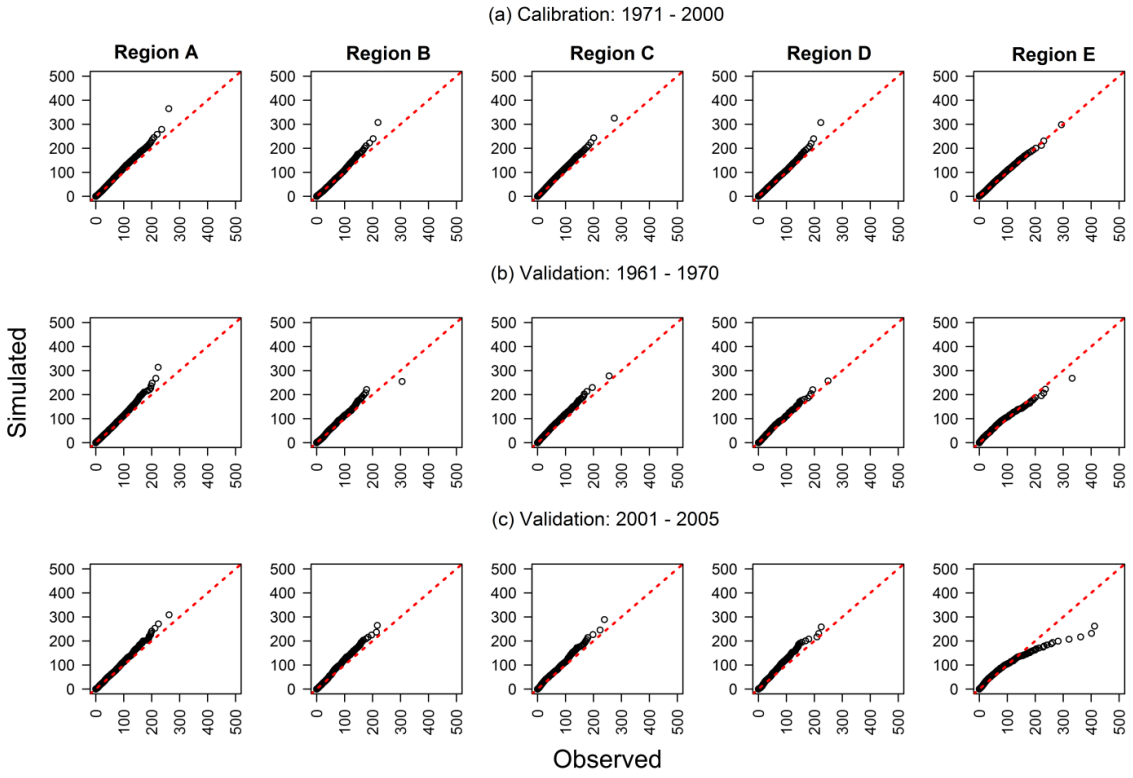

Figure 3-8: Q–Q plots of observed and simulated monthly precipitation totals (in mm) pooled over the number of sites in each region for the calibration (1971–2000) and two validation (1961–1970 and 2001–2005) periods

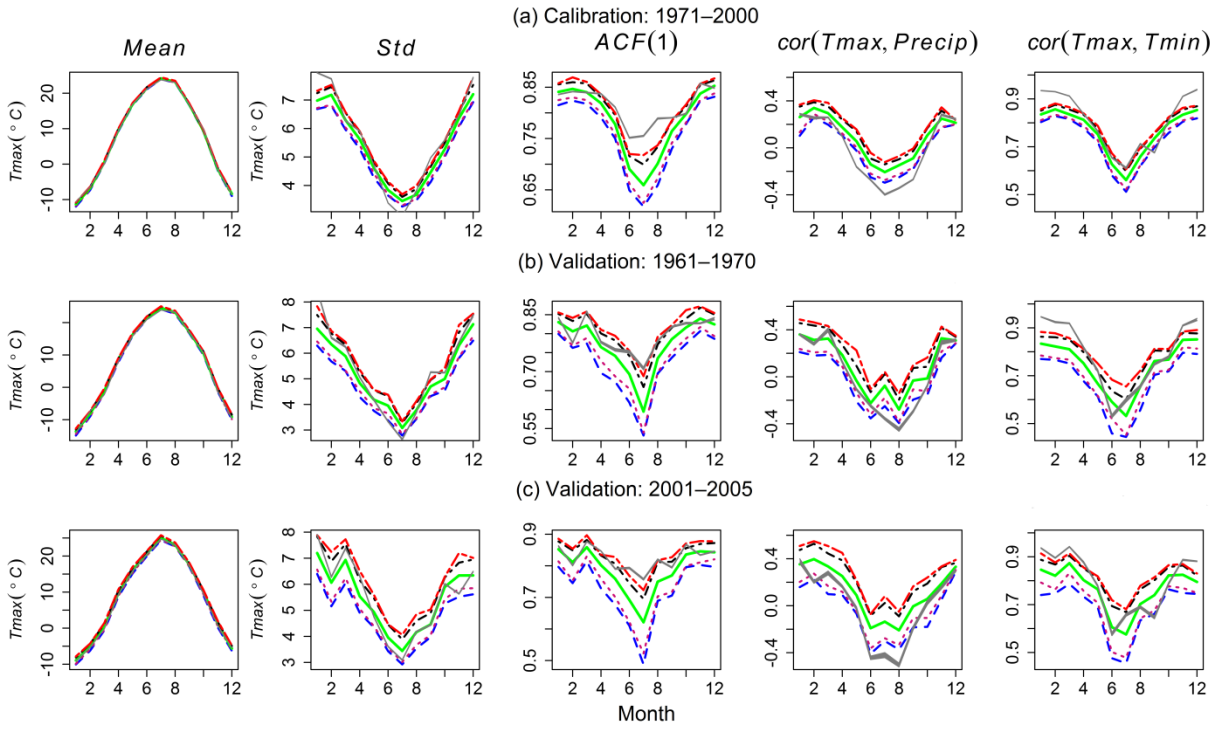

Figure 3-9: Comparison of observed and simulated values of selected statistics of *Tmax* – *Mean*, standard deviation (*Std)*, lag-1 autocorrelation function (*ACF*(1)), and correlation between maximum temperature and precipitation (*cor*(*Tmax,Precip*)) and minimum and maximum temperatures (*cor*(*Tmax,Tmin*)) – for the calibration and two validation periods, together with distributions obtained from 39 imputations of observed data. Remaining convention is the same as in Figure 3-6

Figure 3-11 (left column) shows distributions (i.e. boxplots) of winter and summer seasonal maxima of daily precipitation amounts for Hudson Bay (GG89), Fort McMurray (GG35), Saskatoon (GG77), Edmonton (GGG4) and Medicine Hat (GG20), selected respectively from regions A–E (Table A-1, Appendix A). For each location, the boxplots represent distributions of 95th percentile values derived from 100 simulations, each consisting of one seasonal maximum or minimum per year. The observed values for all locations were well simulated for both seasons. For most cases, the observed value lies within the inter-quartile range of the simulated distribution. For the case of temperature, seasonal maxima of *Tmax* were evaluated to illustrate simulation of summer extremes, while seasonal minima of *Tmin* were evaluated to show simulation of winter extremes. The results of this evaluation are shown in

Figure 3-12 (first column). As for the case of precipitation, it is evident that the GLMs were also able to satisfactorily simulate upper and lower tail behavior of temperature extremes.

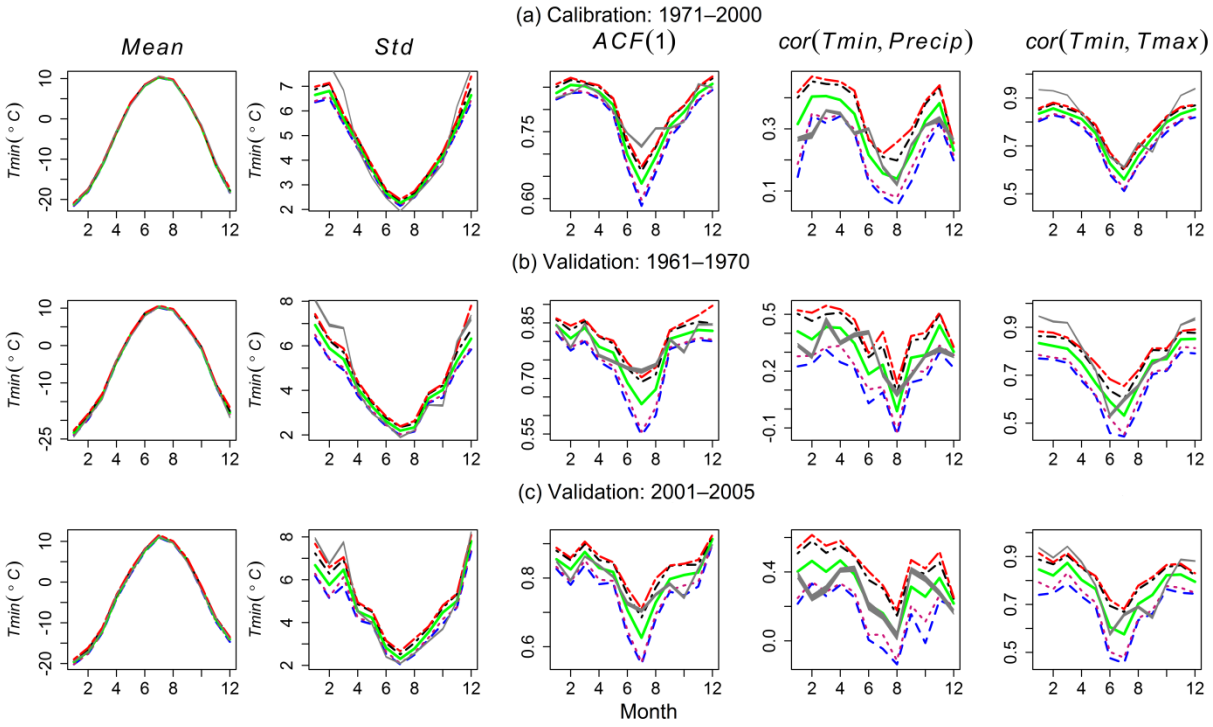

Figure 3-10: Same as in Figure 3-9 but for *Tmin*

For evaluating simulations of selected climate indices (see sub-section 3.3.4.1), we concentrated on the same selected locations as for the precipitation and temperature extremes presented above. Detailed graphical results were omitted for the calibration period, but a summary of the observations made is presented below. For temperature related indices (i.e. *tncw10* and *txhw90*), it can be stated that the GLMs were able to simulate well observed median values for both winter and summer, given that the observed values were within the inter-quartile range of the simulated distribution for most of the cases. For precipitation related indices, a comparison of observed and simulated frequency-based indices (*pxcdd*; *pxcwd*) and mean length of wet/dry spells (*pwsav*; *pdsav*) suggested satisfactory performance of the GLMs. In general,

GLMs performed relatively better in summer than in winter. Overall, temperature related indices were better reproduced than the precipitation related indices.

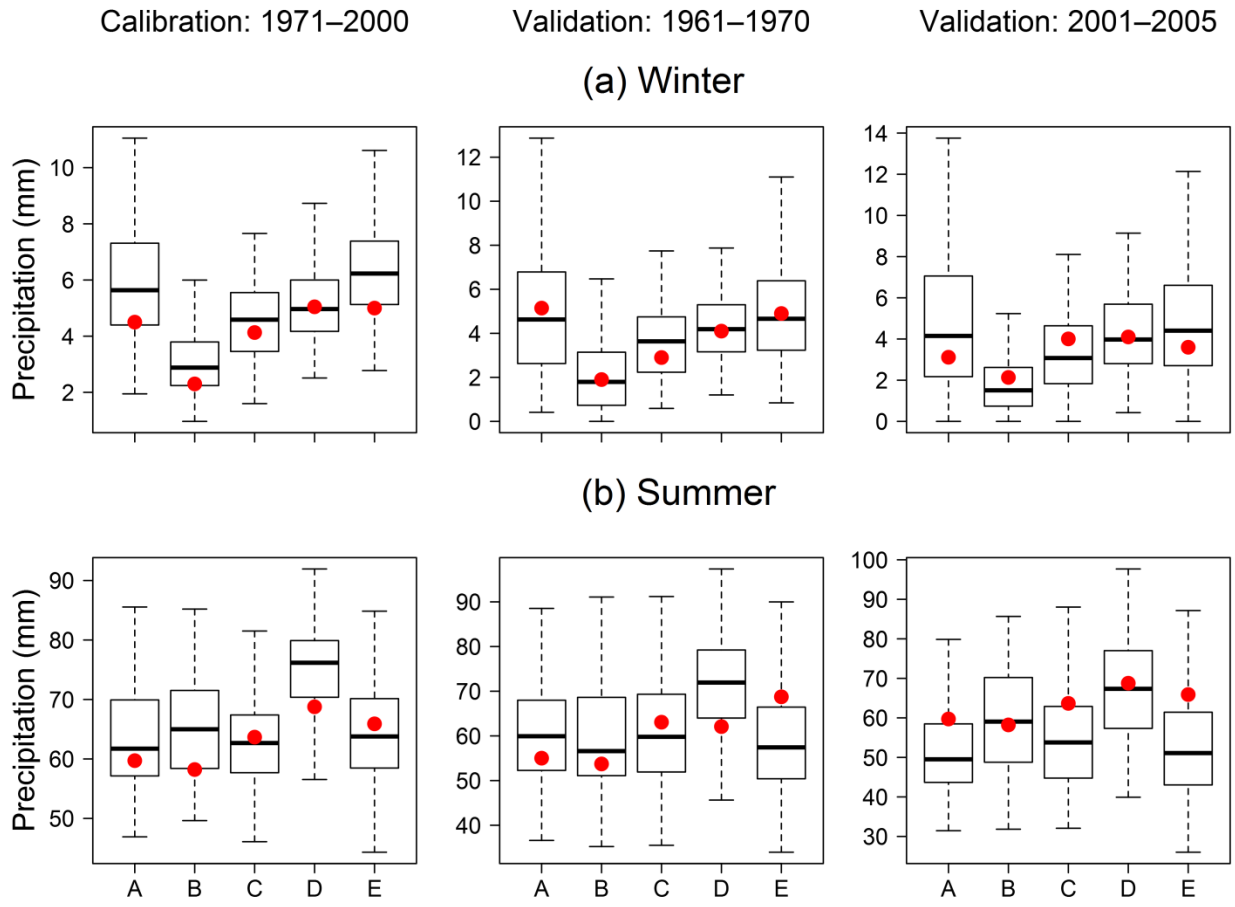

Figure 3-11: Evaluation of GLM performance for simulating (a) winter and (b) summer extremes of precipitation amounts for the calibration (1971–2000) and two validation (1961–1970 and 2001–2005) periods for Hudson Bay, Fort McMurray, Saskatoon, Edmonton, and Medicine Hat, located respectively in each of the five homogeneous regions A–E. In each boxplot, the box corresponds to the interquartile range, the line in the middle of the box to the median value and the whiskers to either the maximum or minimum value of the simulated distribution. For each season and time period, red dots indicate the median of observed annual maximum daily values. Boxplots represent distributions of 95th percentile values derived from 100 simulations, each consisting of one seasonal maximum per year

3.4.2 Validation of GLMs

The calibrated models were evaluated by generating 100 realizations of daily precipitation and temperature fields for the pre- and post-calibration periods (i.e. 1961–1970 and 2001–2005). In summary, for both validation periods, *ACF*(1), $Mean^{cond}$, and Std^{cond} of simulated precipitation

sequences were satisfactorily reproduced except P_W, which was overestimated by the models. Figures are omitted as the results were very similar to those for the calibration period. Figure 3-8 (b and c) shows a comparison of observed and simulated monthly precipitation totals. On the regional level, monthly precipitation totals were well reproduced for most of the regions except region E, where the observed values were underestimated for the 2001–2005 period (Figure 3-8c). For this region, which corresponds to the Rocky Mountains, the spatial structure of precipitation is probably more complex than in other regions of the study area. Jiang (2003) noted that modelling of precipitation in mountainous areas is particularly challenging because of the multiscale nature of the complex terrain, interactions between terrain and airflow, the complex role of latent heating/cooling, and the complexity of cloud physics. The results of comparison of observed and simulated statistics of *Tmax* and *Tmin* are shown in Figure 3-9 (b and c) and Figure 3-10 (b and c), respectively. These results suggested that the model performance is very similar to that discussed for the calibration period.

Next, the evaluation of the GLM framework in reproducing seasonal precipitation and temperature extremes is discussed. For the case of precipitation extremes, it is evident from column two and three of Figure 3-11 that the winter and summer extremes were satisfactorily reproduced. The results of temperature related extremes are shown in Figure 3-12 (column two and three) for winter and summer seasons. Again, the GLMs were able to capture both lower and upper tail behavior of the observed distributions for the two validation periods. As noted for the calibration case, temperature related extremes are better reproduced than the precipitation related extremes.

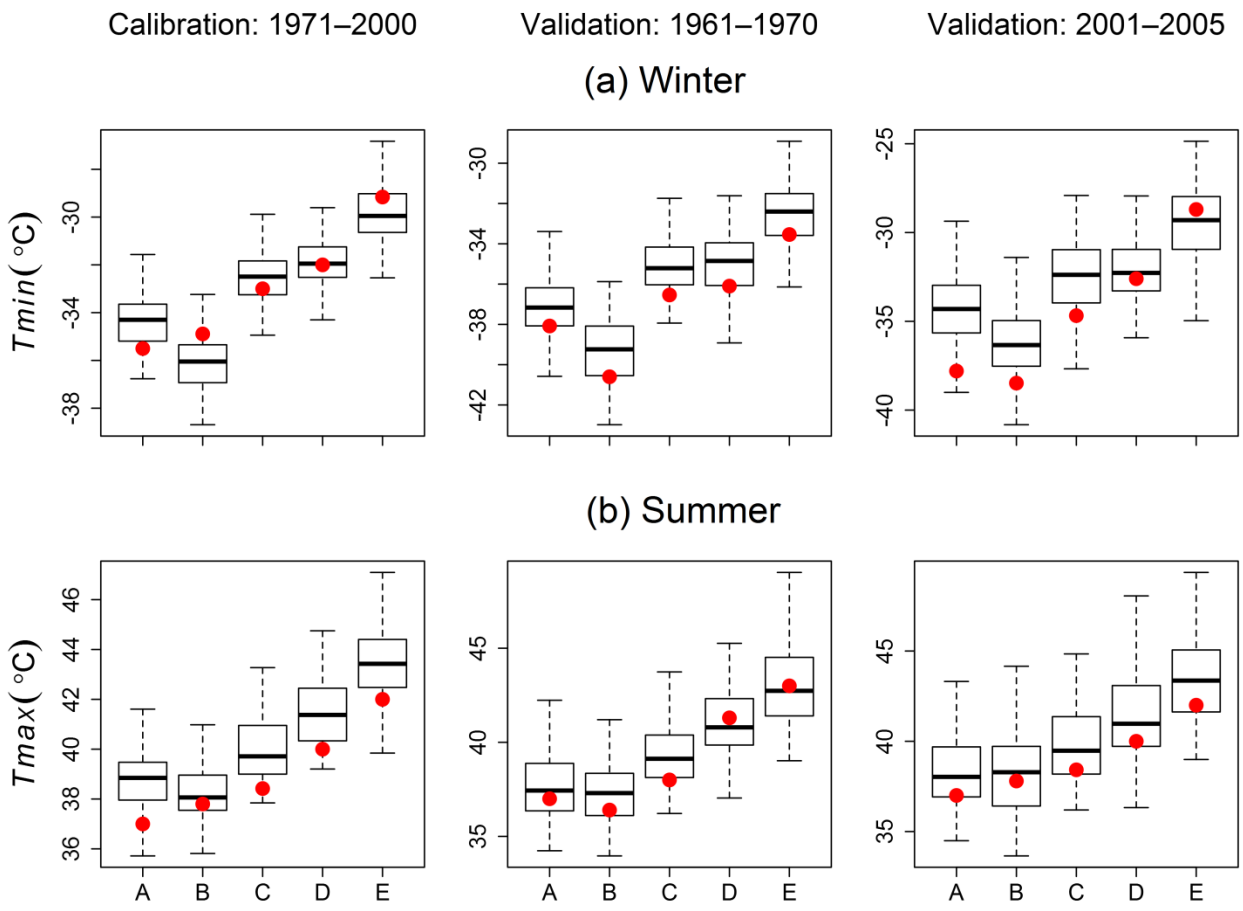

Figure 3-12: Evaluation of GLM performance for simulating extreme values of (a) winter *Tmin* and (b) summer *Tmax* temperatures for the calibration (1971–2000) and two validation periods (1961–1970 and 2001–2005) for Hudson Bay, Fort McMurray, Saskatoon, Edmonton, and Medicine Hat, located respectively in each of the five homogeneous regions A–E. Boxplots represent distributions of 95th percentile values derived from 100 simulations, each consisting of one seasonal maximum in case of *Tmax* and minima in case of *Tmin* per year. Remaining convention is the same as in Figure 3-11

Figure 3-13 shows boxplots of selected climate indices for winter and summer seasons for the 1961–1970 period only. For each location, the boxplot represents distributions of indices derived from 100 realizations. For temperature related indices (i.e. *tncw10* and *txhw90*), observed values lie within the inter-quartile range of the simulated values for most cases. For the case of precipitation, frequency-based indices (i.e. *pxcdd* and *pxcwd*) and mean length of wet/dry spells (*pwsav*; *pdsav*) were satisfactorily captured in both seasons. The performance of the GLMs was generally better in summer than in winter. Similar results were realized for the other validation

period (2001–2005) for most of the regions, except region E, as discussed above. A probable explanation for these discrepancies could be that this region experienced severe drought conditions in 2001–2005 period compared to the 1971–2000 period used for calibration.

3.5 Summary and Conclusions

The main goal of this study was to explore the suitability of GLMs for modelling multisite precipitation and temperature sequences in the Canadian Prairie Provinces using large-scale atmospheric fields from NCEP reanalysis-I and the PDO and PNA as exogenous covariates. The logistic regression approach was used to model precipitation occurrences, while the two-parameter gamma distribution was used to model precipitation amounts. A jointly fitted model comprising the mean and dispersion components was used to model daily minimum and maximum temperatures. The suitability of the fitted GLMs for characterizing precipitation and temperature fields in terms of (a) simulating their mean values at the daily, monthly and seasonal scales, (b) characteristics of extreme values, (c) intervariable relationships and (d) selected climate indices were investigated using independent observations from pre- and post-calibration periods. The following conclusions can be drawn from the various analyses presented:

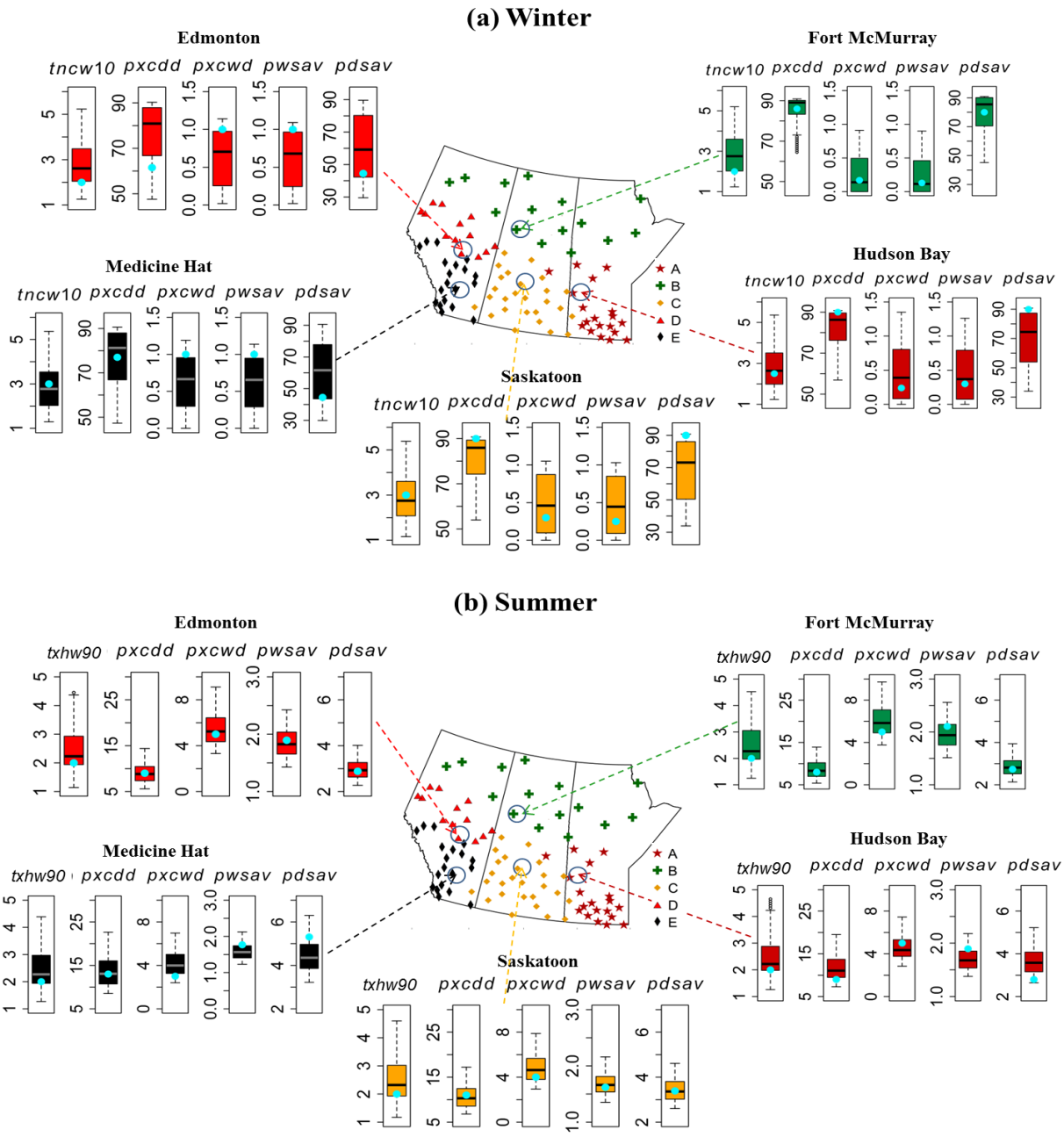

Figure 3-13: Evaluation of GLM performance for simulating climate indices in (a) winter and (b) summer for the 1961–1970 validation period for Hudson Bay, Fort McMurray, Saskatoon, Edmonton, and Medicine Hat, located respectively in each of the five homogeneous regions A–E. The remaining convention is the same as in Figure 3-11

(1) Based on residual analysis, it was found that a single model for precipitation sequences could not be justified for the entire study area. Therefore, separate models were developed on the basis of five pre-defined homogeneous regions covering the study area. Following

this approach, residual plots for each region showed significant improvement in the performance of the fitted GLMs.

(2) For both calibration and validation periods, there was generally good agreement between the simulated and observed values of various precipitation and temperature characteristics for each month of the year. Most of the statistical features are generally well reproduced, except the proportion of wet days, which was slightly overestimated. The observed characteristics lie generally within the 2.5^{th} and 97.5^{th} percentiles of the simulated values. The uncertainty bands due to missing observed values are found to be quite large, especially for the winter season. In general, the simulated values of precipitation characteristics were more variable than those of temperature fields.

(3) The fitted GLMs were able to capture spatial and inter-variable dependence structure. Distance-based inter-site correlations were well reproduced by the GLMs. The temporal correlations between precipitation and *Tmin* were well captured by the models. However, the temporal dependence between summer precipitation and *Tmax* was generally underestimated. This could result from the inability of the GLMs to capture short duration, localized and generally more intense convective precipitation storms often found in summer in the study region.

(4) The fitted models were also assessed for robustness in terms of their ability to reproduce characteristics of extreme events and some of the commonly used climate indices. In summer, the performance of the models was generally better than in winter as the observed values for most indices and 95^{th} percentiles of the winter and summer seasonal extremes fell mostly within the inter-quartile range of the simulated values. Overall, hot and cold temperature related indices and characteristics of temperature extremes were better

reproduced than the precipitation related indices and characteristics of extreme precipitation amounts.

Finally, it can be concluded that apart from few limitations (such as overestimation of proportion of wet days), the GLM framework has the potential for multisite multivariate modelling of daily precipitation and temperature fields. This framework is able to describe satisfactorily mean and extreme climate characteristics using NCEP reanalysis-I predictors and teleconnection indices. So far, we have not come across any plausible weather generator that can reasonably be applied to a huge and clearly inhomogeneous region studied in this paper. The next phase of this study is to use the fitted models for downscaling climate projections from state-of-the-art GCMs participating in the Coupled Model Intercomparison Project Phase 5 (CMIP5) of the Intergovernmental Panel on Climate Change. Such analyses will furnish additional opportunities for further evaluation of the GLM framework, in particular, validity of the key assumptions of statistical downscaling, including temporal invariance, discussed in Wilby et al. (2004).

Acknowledgements

The financial support from the Global Institute for Water Security and School of Environment and Sustainability is gratefully acknowledged. Thanks are due to Eva Mekis from Environment Canada for providing access to adjusted precipitation and temperature data used in this study. We also thank Yanping Li for shedding light on meso-scale meteorological processes in the Canadian Prairie Provinces and Sun Chun for the useful comments on an earlier version of this paper. The invaluable programming assistance of Gonzalo Sapriza Azuri is much appreciated. The indices of extreme events were computed using the STARDEX project FORTRAN routines. Finally, we thank Richard Chandler from University College London and an anonymous referee for very detailed and useful comments which helped improve the quality of the analyses presented in the paper.

References

Ambrosino C, Chandler RE, Todd MC (2014), Rainfall-derived growing season characteristics for agricultural impact assessments in South Africa, *Theoretical and Applied Climatology,* 115,411 – 426.

Apipattanavis, S., G. Podesta´, B. Rajagopalan, and RW Katz (2007), A semiparametric multivariate and multisite weather generator, *Water Resource Research,* 43, W11401, doi:10.1029/2006WR005714.

Asong ZE, Khaliq MN, Wheater HS (2015), Regionalization of precipitation characteristics in the Canadian Prairie Provinces using large-scale atmospheric covariates and geophysical attributes, *Stochastic Environmental Research and Risk Assessment*, 29 (3), 875-892.

Bergin E, Buytaert W, Onof C, Wheater HS (2012), Downscaling of rainfall in Peru using generalised linear models. In World Congress on Water, Climate and Energy. Dublin, Ireland, 13–18 May.

Bonsal BR, Aider R, Gachon P, Lapp S (2012), An assessment of Canadian prairie drought: past, present, and future, *Climate Dynamics,* 41, 501–516. doi: 10.1007/s00382-012-1422-0.

Borchert JA (1950), The Climate of the Central North American grassland, *Annals of the Association of American geographers,* 40, 1–39.

Buishand TA (1978), Some remarks on the use of daily rainfall models, *Journal of Hydrology,* 36, 295 – 308.

Cavazos T, Hewitson BC (2005), Performance of NCEP variables in statistical downscaling of daily precipitation, *Climate Research,* 28, 95–107.

Chandler RE (2005), On the use of generalized linear models for interpreting climate variability, *Environmetrics*, 16(7), 699–715.

Chandler RE (2014), Rglimclim: A multisite, multivariate weather generator based on generalized linear models. R package version 1.3-0. http://homepages.ucl.ac.uk/~ucakarc/work/glimclim.html.

Chandler RE, Bate SM (2007), Inference for clustered data using the independence log-likelihood, *Biometrika* 94(1), 167–183.

Chandler RE, Scott EM (2011), Statistical methods for trend detection and analysis in the environmental sciences. Wiley, Chichester.

Chandler RE, Wheater HS (2002), Analysis of rainfall variability using generalized linear models-a case study from the West of Ireland, *Water Resources Research*, 38 (10), 1192. doi:10.1029/2001WR000906.

Chen, J., F. Brissette, R. Leconte, A. Caron (2012), A versatile weather generator for daily precipitation and temperature, *Transactions of the American Society of Agricultural and Biological Engineers*, 55(3), 895-906.

Chun KP, Wheater HS, Nazemi A, Khaliq MN (2013), Precipitation downscaling in Canadian Prairie Provinces using the LARS-WG and GLM approaches, *Canadian Water Resources Journal*, 38 (4), 311-332.

Coe R, Stern RD (1982), Fitting models to daily rainfall, *Journal of Applied Meteorology*, 21, 1024–1031.

D'Onofrio, D., Palazzi, E., von Hardenberg, J., Provenzale, A., Calmanti, S (2014), Stochastic Rainfall Downscaling of Climate Models, *Journal of Hydrometeorology*, 15(2), 830–843.

Dibike YB, Gachon P, St-Hilaire A, Ouarda TBMJ, Nguyen VT (2008), Uncertainty analysis of statistically downscaled temperature and precipitation regimes in Northern Canada, *Theoretical and Applied Climatology*, 91, 149–170.

Fowler HJ, Blenkinsopa S, Tebaldi C (2007), Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, *International Journal of Climatology*, 27, 1547–1578.

Frost AJ, Charles SP, Timbal B, Chiew FHS, Mehrotra R, Nguyen KC, Chandler RE, McGregor JL, Fu G, Kirono DGC, Fernandez E, Kent DM (2011), A comparison of multi-site daily rainfall downscaling techniques under Australian conditions, *Journal of Hydrology*, 408, 1–18.

Furrer EM, Katz RW (2007), Generalized linear modeling approach to stochastic weather generators, *Climate Research*, 34, 129–144.

Giorgi F (2006), Regional climate modeling: status and perspectives, *Journal of Physics*, IV 139, 101-118.

Goodess CM (2003), Statistical and regional dynamical downscaling of extremes for European regions: STARDEX, *Eggs*, 6, 25– 29.

Gutiérrez, J. M., San-Martin, D., Brands S, Manzanas R., Herrera S (2013), Reassessing Statistical Downscaling Techniques for their Robust Application under Climate Change Conditions, *Journal of Climate*, 26, 171–188.

Haylock MR, Cawley GC, Harpham C, Wilby RL, Goodess CM (2006), Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios, *International Journal of Climatology*, 26, 1397–1415.

Hogg, E. H., D. T. Price, and T. A. Black (2000), Postulated feedbacks of deciduous forest phenology on seasonal climate patterns in the Western Canadian Interior, *Journal of Climate*, 13, 4229–4243.

Hundecha Y, Pahlow M, Schumann A (2009), Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes, *Water Resource Research* 45, W12412. doi:10.1029/2008WR007453.

Huth R (2002), Statistical downscaling of daily temperature in Central Europe, *Journal of Climate*, 15(13), 1731–1742.

Jeong, D. I., St-Hilaire, A., Ouarda, T. B., and Gachon, P. (2012), Multisite statistical downscaling model for daily precipitation combined by multivariate multiple linear regression and stochastic weather generator, *Climatic Change*, 114(3-4), 567- 591.

Jiang Q (2003), Moist dynamics and orographic precipitation, *Tellus A*, 55, 301–316.

Kalnay E et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bulletin of American Meteorological Society*, 77(3), 437 – 471.

Katz RW (1977), Precipitation as a chain-dependent process, *Journal of Applied Meteorology*, 16, 671 – 676.

Katz RW, Parlange MB, Tebaldi C (2003), Stochastic modeling of the effects of large scale circulation on daily weather in the southeastern U.S., *Climatic Change*, 60,189–216.

Kenabatho P, McIntyre N, Chandler R, Wheater H (2012), Stochastic simulation of rainfall in the semi-arid Limpopo basin, Botswana, *International Journal of Climatology*, 32,1113–1127.

Khalili M, Nguyen VT, Gachon P (2013), A statistical approach to multi-site multivariate downscaling of daily extreme temperature series, *International Journal of Climatology*, 33, 15–32.

Khaliq MN, L Sushama, A Monette, Wheater H (2014), Seasonal and extreme precipitation characteristics for the watersheds of the Canadian Prairie Provinces as simulated by the NARCCAP multi-RCM ensemble, *Climate Dynamics*, doi: 10.1007/s00382-014-2235-0.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Reviews of Geophysics*, 48, RG3003, doi:10.1029/2009RG000314.

McCullagh P, Nelder JA (1989), *Generalized Linear Models*, second edition, Chapman and Hall, London

Mehrotra, R., Sharma, A (2007), A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability, *Journal of Hydrology*, 335(1-2), 180-193.

Mekis É, Vincent LA (2011), An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada, *Atmosphere-Ocean*, 49(2), 163–177.

Mladjic B, Sushama L, Khaliq MN, Laprise R, Caya D, Roy R (2011), Canadian RCM projected changes to extreme precipitation characteristics over Canada, *Journal of Climate*, 24, 2565–2584.

Philips D (1990), Climates of Canada. Canadian Government Publishing Ottawa, Ontario

R Core Team (2014), R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL http://www.R-project.org/.

Rajagopalan, B., U. Lall (1999), A k-nearest-neighbor simulator for daily precipitation and other variables, *Water Resource Research*, 35(10), 3089 – 3101.

Richardson CW (1981), Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resource Research*, 17,182–190.

Richardson CW, Wright DA (1984), WGEN: A Model for Generating Daily Weather Variables. U. S. Department of Agriculture, *Agricultural Research Service*, ARS-8, 83p.

Schoof JT, Pryor SC (2001), Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks, *International Journal of Climatology*, 21, 773–790.

Steinschneider, S., C. Brown (2013), A semiparametric multivariate, multisite weather generator with low-frequency variability for use in climate risk assessments, *Water Resource Research*, 49, 7205–7220.

Semenov MA, P. Stratonovitch (2010), Use of MultiModel Ensembles from Global Climate Models for Assessment of Climate Change Impacts, *Climate Research*, 41, 1–14.

Stern RD, Coe R (1984), A model fitting analysis of daily rainfall data, *Journal of the Royal Statistical Society,* Series A, 147, 1 – 34.

Todorovic P, Woolhiser DA (1975), A stochastic model of n-day precipitation, *Journal of Applied Meteorology,* 14, 17 – 24.

Turco M, Quintana-Segui P, Llasat, MC, Herrera S, Gutierrez, JM (2011), Testing MOS precipitation downscaling for ENSEMBLES regional climate models over Spain, *Journal of Geophysical Research,* 116, 1 -14.

Vincent LA, Milewska EJ, Hopkinson R, Malone L (2009), Bias in minimum temperature introduced by a redefinition of the climatological day at the Canadian synoptic stations, *Journal of Applied Meteorology and Climatology,* 48, 2160–2168.

Wheater HS, and Gober P (2013), Water Security in the Canadian prairies: science and management challenges, *Philosophical Transactions of the Royal Society A,* 371, 20120409. doi: 10.1098/rsta.2012.0409.

Wilby RL, Wigley TML (1997), Downscaling general circulation model output: a review of methods and limitations, *Progress in Physical Geography,* 21,530–548.

Wilby RL, Wigley TML (2000), Precipitation predictors for downscaling: Observed and general circulation model relationships, *International Journal of Climatology,* 20, 641-661.

Wilby, RL., CW., Dawson, and EM, Barrow (2002), SDSM—a decision support tool for the assessment of regional climate change impacts, *Environmental Modelling & Software,* 17 (2), 145-157.

Wilks DS (1998), Multisite generalization of a daily stochastic precipitation generation model, *Journal of Hydrology,* 210, 178–191.

Yang C, Chandler RE, Isham VS, Wheater HS (2005), Spatial-temporal rainfall simulation using generalized linear models, *Water Resource Research,* 41,1–13.

Zorita E, von Storch H (1997), A survey of statistical downscaling techniques, GKSS report 97/E/20, *GKSS Research Center,* Geesthacht

CHAPTER 4
PROJECTED CHANGES IN PRECIPITATION AND TEMPERATURE CHARACTERISTICS OVER THE CANADIAN PRAIRIE PROVINCES USING THE GENERALIZED LINEAR MODEL MULTISITE MULTIVARIATE STATISTICAL DOWNSCALING APPROACH

Many recent studies have shown that changes in the global climate since the 20th century were mostly due to the anthropogenic GHG emissions than the natural variability of the climate. Rising GHG concentrations in the atmosphere tend to heighten the natural greenhouse effect resulting in an imbalance in the earth's radiative energy budget, thus, human induced climate change. Rising global temperatures, sea level rise due to melting glaciers, changes in the precipitation patterns and variability which cause floods and droughts are some of the impacts of the changing climate. As the increased variability in the mean and extreme weather events directly affect socio-economic and environmental sectors, projections of changes in the characteristics of these events are important to inform climate change mitigation- and adaptation-related decision making.

Until now the most reliable source of climate change information is from Atmosphere Ocean General Climate Models (AOGCMs). The output of these models is available at very coarse spatial resolution not readily suitable for hydrological impact studies at the station scale. Furthermore, AOGCM output is associated with various sources of uncertainties, and also has limited or no ability to capture sub-grid scale processes which are relevant for many environmental and water resources protection studies. To bridge this gap, downscaling methods (i.e. statistical and dynamical downscaling) have often been utilized to transform AOGCM information to local and regional scale resolution. The AOGCMs considered for this study from the CMIP5 ensemble are also called Earth System Models. The research presented in this chapter was motivated by a desire to explore the suitability of the GLM framework in downscaling

future climate projections provided by AOGCMs to station scale meteorological variables. This chapter contains the following manuscript:

1. **Asong ZE**, Khaliq MN, Wheater HS (2015), Projected changes in precipitation and temperature characteristics over the Canadian Prairie Provinces using the Generalized Linear Model multisite multivariate statistical downscaling approach. Submitted to *Journal of Hydrology, Paper # HYDROL20972* (status: "under review" as of 14/12/2015).

Abstract

Atmosphere-Ocean General Circulation Models (AOGCMs) are important tools for simulating future climate that might result from anthropogenic modification of the atmosphere. However, outputs from these models cannot reliably be applied directly in many environmental and water resources related studies because of coarse spatial resolution and limitations in representing sub-grid scale processes. To bridge this gap, downscaling methods (both statistical and dynamical) are utilized to transform AOGCM outputs to local- and regional-scale resolutions. In this study, a multisite multivariate statistical downscaling approach based on the Generalized Linear Model (GLM) framework is developed to downscale daily observations of precipitation and minimum and maximum temperatures from 120 sites located across the Canadian Prairie Provinces: Alberta, Saskatchewan and Manitoba. First, large scale atmospheric covariates from the National Center for Environmental Prediction (NCEP) Reanalysis-I, teleconnection indices, geographical site attributes, and observed precipitation and temperature records are used to calibrate GLMs for the 1971–2000 period. Then the calibrated models are used to generate daily sequences of precipitation and temperature for the 1962–2005 historical (conditioned on NCEP predictors), and future period (2006–2100) using outputs from five CMIP5 (Coupled Model Intercomparison Project Phase-5) Earth System Models corresponding to Representative Concentration Pathway (RCP): RCP2.6, RCP4.5, and RCP8.5 scenarios. The

results indicate that the fitted GLMs are able to capture spatiotemporal characteristics of observed precipitation and temperature fields. According to the downscaled future climate, mean precipitation is projected to increase in summer and decrease in winter while minimum temperature is expected to warm faster than the maximum temperature. Climate extremes are projected to intensify with increased radiative forcing.

Keywords: GLMs; extreme events; precipitation; temperature; AOGCMs; CMIP5; projected changes

4.1 Introduction

The Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (IPCC, 2013) demonstrates that the global climate is warming and most of the observed changes are likely due to increases in the concentration of anthropogenic greenhouse gases (GHGs). The impacts of a warming climate on global and regional water resources are many and have been re-emphasized recently (e.g., Arnell and Lloyd-Hughes, 2014; Wheater and Gober, 2015). The intensity, frequency and magnitude of extreme climate events such as extreme precipitation, droughts, heat waves and cold waves associated with global warming are likely to change over space and time (Kharin et al., 2013; Fischer and Knutti, 2014; Jeong et al., 2014, 2015). As the increased variability in the mean and extreme climate events directly affects socio-economic and environmental sectors, projections of changes in the characteristics of these events are important to inform climate change mitigation- and adaptation-related decision making.

Atmosphere-Ocean General Circulation Models (AOGCMs) are regarded as the most credible tools to simulate time series of climatic variables at the global scale, accounting for the effects of rising GHGs in the atmosphere. However, there exist many fundamental challenges in the application of AOGCM outputs at local scales primarily due to the coarse resolution which is typically of the order of a few hundred kilometres (Fowler et al., 2007; Maraun et al., 2010).

Consequently, these models are unable to resolve important sub-grid scale features such as topography, clouds, and land use which influence climate variability at local to regional scales (Xu, 1999). Therefore, the outputs of AOGCMs cannot at the moment be directly used for local scale climate impact investigations which generally require climatic data at certain points of interest in space (Bates et al., 1998). To overcome these limitations, downscaling techniques have been developed to transform the AOGCM outputs to fine resolution information required for impact assessment at local scales. Downscaling methods have been classified into two broad groups (i.e., dynamical and statistical) (Wilby and Wigley, 1997).

The dynamical downscaling techniques use Regional Climate Models (RCMs) to predict finer scale regional climate variables using boundary conditions from an AOGCM (Frei et al., 2006). Although dynamical downscaling is capable of producing spatially distributed climatic predictions over smaller regions of interest (e.g., Wang et al., 2015), direct application of outputs from RCMs for climate change impact studies is still limited, perhaps due to their complicated design, high computational cost and/or bias partly originating from the driving AOGCM. Compared to the dynamical downscaling approach, statistical downscaling (SD) is based on establishing empirical relationships between large scale predictors which are assumed to be skilfully simulated by the AOGCMs and the target local scale variables (predictands) under the so-called *perfect prognosis* (PP) approach. Using this approach, reanalysis outputs for a representative historical period are used as predictors while simultaneous local scale observations are used as predictands for model calibration. The optimally calibrated model obtained using these observed data is applied to the output of different AOGCM scenario runs to obtain future projections of climate change.

In general, statistical downscaling approaches have been grouped under three broad categories (Wilby et al., 2004): regression-based methods, weather typing approaches, and

stochastic weather generators (WGs). Individual downscaling strategies differ according to the choice of predictor variables, the statistical fitting procedure and most importantly the transfer function used. In regression-based methods, linear or non-linear quantitative relationships between AOGCM outputs (predictors) and local scale variables (e.g., precipitation, temperature) are derived. Examples include Generalized Linear Models (GLMs) (Chandler and Wheater, 2002); multiple linear regression (Sachindra et al., 2012); canonical correlation analysis (von Storch and Zwiers, 1999); artificial neural networks (Hewitson and Crane, 1996); and support vector machines (Chen et al., 2010).

Weather typing approaches aim at classifying the large scale weather patterns into a number of discrete states based on their synoptic resemblance, and links them with station scale observations (Goodess and Palutikof, 1998). Subsequently, corresponding to the future states of the large scale weather characterized by the AOGCMs, the station scale weather pertaining to the future is derived. These derivations can be made either by resampling from the observed distribution of the variable conditioned on the circulation patterns produced by AOGCMs, or by first generating synthetic sequences of the large scale weather fields using for example, Monte Carlo techniques and then resampling from the simulated data. The mean or frequency distribution of the local scale climate is then computed by weighting the local climate states with the relative frequencies of the large scale weather classes. Meteorological analogs (Charles et al., 2013) and recursive partitioning (Schnur and Lettenmaier, 1998) are typical examples of the weather classification approaches. The main benefit of weather typing techniques is that they are capable of downscaling highly non-linear relationships between the predictors and the predictand(s) of interest. However, these techniques are based on the assumption that a certain large scale weather pattern realized in the past will lead to the same station scale weather condition in the future. Another issue is that if a weather type simulated by the AOGCM for the

future is not found among the historical weather types, determination of the station weather becomes chaotic.

WGs are statistical models which are capable of producing synthetic weather sequences which capture most essential features of the observed weather (Wilks, 1998). For obtaining the weather pertaining to the future, the parameters of the WG are adjusted according to changes predicted by the AOGCM. The most relevant strength of WGs is that they can be used to generate time series of weather data of any length and number. There are two fundamental types of daily WGs, based on the approach to model daily precipitation occurrence: the Markov chain and spell-length approach. In some studies, precipitation has been modelled using a two-stage process involving separate models for precipitation occurrence and amounts when wet (Todorovic and Woolhiser, 1975; Stern and Coe, 1984). Daily precipitation occurrence is often modeled using a two-state Markov process corresponding to wet and dry states (e.g., Richardson, 1981; Katz et al., 2003), while the gamma distribution has commonly been used to model precipitation amounts (e.g., Katz, 1977). GLMs (McCullagh and Nelder, 1989; Chandler and Wheater, 2002; Furrer and Katz, 2007) offer a framework that unifies and extends many of the existing approaches that have been proposed to model precipitation occurrences. According to Maraun et al. (2010), GLM-based WGs compare favourably with other existing state-of-the-art techniques. For example, Frost et al. (2011) evaluated relative strengths of six multi-site daily rainfall modelling and downscaling techniques and found GLMs to reproduce the occurrence and amounts statistics satisfactorily.

Most weather generators concentrate on individual sites (e.g., Rajagopalan and Lall, 1999) and therefore are unable to represent the spatial structure of the observed climatic variables. For many water resources design projects, particularly in large river basins, it is important to model simultaneous sequences of multiple variables (e.g., precipitation and temperature) over large

heterogeneous areas, while maintaining physically plausible spatial, temporal and intervariable relationships. The GLM-based Rglimclim WG (Chandler, 2014) provides a flexible framework for accomplishing such challenging tasks.

The research presented in this study was motivated by a desire to explore the suitability of the GLM framework for downscaling future climate projections. The National Center for Environmental Prediction (NCEP) Reanalysis-I dataset (Kalnay et al., 1996) was used to train the GLM framework and the AOGCM outputs were used for projecting future precipitation and temperature sequences over the Alberta, Saskatchewan and Manitoba provinces of Canada, a region consisting of 47 diverse watersheds including the Saskatchewan, Athabasca, Peace, and Churchill River basins. This region was hard-hit by recurrent severe droughts in 1988, and 1999–2005, and floods in 2010, 2011, 2013, and 2014. Given the projected global changes in climate for the 21st century, a deeper understanding of the impacts of climate change on local to catchment and regional scales is important for regional water resources management related decision making and planning for future droughts and floods.

This paper is organized as follows: Section 4.2 provides a brief description of the study area and the data used in the study. Section 4.3 describes the generic methodology for downscaling AOGCM outputs to station scales. Results of the study are presented and discussed in Section 4.4, while a summary and conclusions are provided in Section 4.5.

4.2 Data

The datasets used in this study included daily total precipitation and minimum (Tmin) and maximum (Tmax) temperatures for the 1961 to 2005 period from a network of 120 stations (Figure 4-1 and Table A-1), obtained from Environment Canada (http://www.ec.gc.ca). Temperature is recorded at 96 of the 120 stations. The underlying daily dataset is categorized as the *Second Generation of Daily Adjusted Precipitation and Temperature Data for Canada*.

These datasets have been quality controlled and adjusted to account for known changes in measurement practices (Mekis and Vincent, 2011).

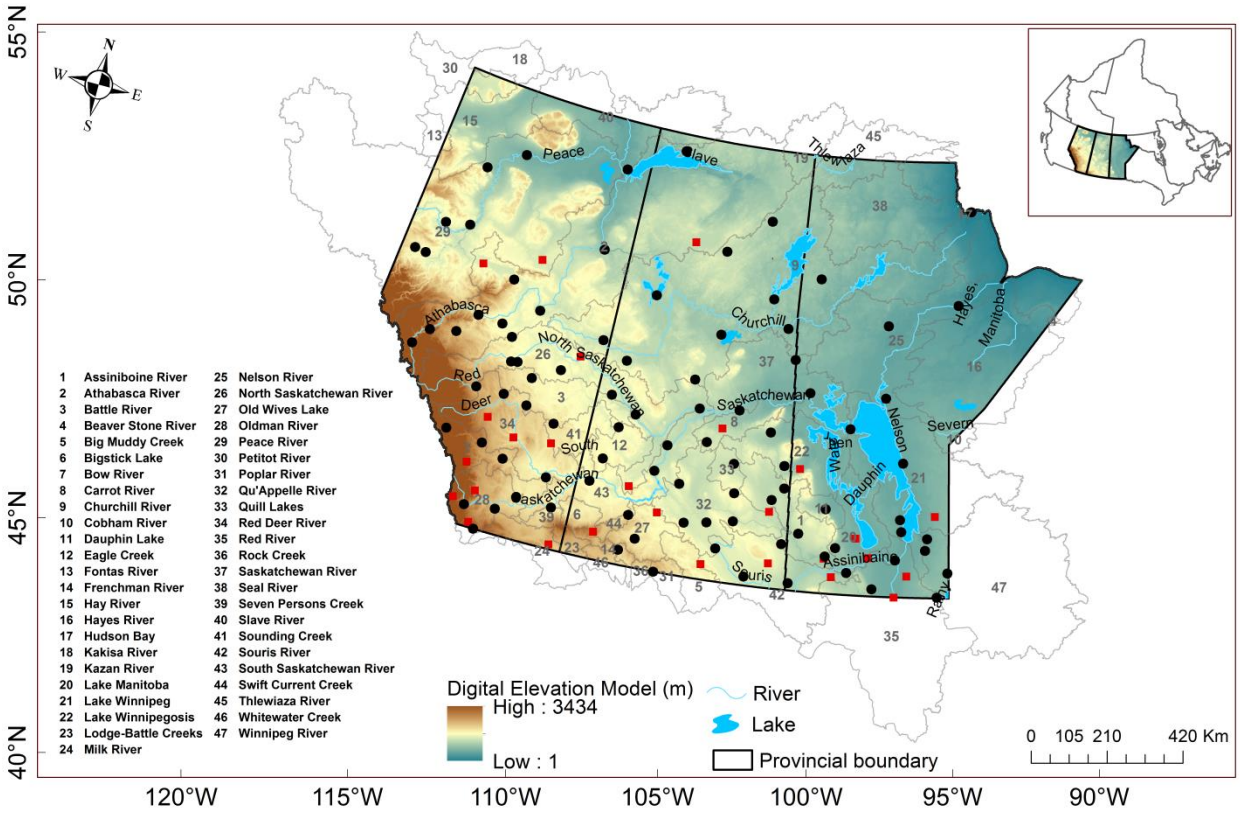

Figure 4-1: Study area and location of observation stations (black dots and red squares). Precipitation is observed at all stations, while temperature is recorded only at stations shown in black. Forty seven watersheds spanning the study area alongside the various codes and names are also shown on the map.

The AOGCM datasets included spatially averaged daily outputs from six CMIP5 multi-model climate change experiments (Table 4-1) over a spatial domain encompassing latitudes 40^{o}N to 70^{o}N and longitudes 130^{o}W to 70^{o}W (Figure 4-2). The choice of domain size and large scale predictors in the statistical downscaling process was based on a number of criteria. Foremost, the predictors should have a strong physical connection to the local weather processes [e.g., Asong et al., 2015] and be reliably simulated by the AOGCMs (Taylor et al., 2012). It is important to note that only the first output per AOGCM corresponding to Representative Concentration Pathway (RCP) RCP2.6, RCP4.5, and RCP8.5 scenarios was downscaled in this

study. The CMIP5 model outputs are available from the archives of the Program for Climate Model Diagnosis and Intercomparison (PCMDI, http://www-pcmdi.llnl.gov) and the Earth System Grid (ESG) data distribution portal (http://www.earthsystemgrid.org). As the AOGCMs have different horizontal resolutions from that of NCEP (Figure 4-2), the extracted AOGCM outputs were re-gridded (via bilinear interpolation) onto the NCEP grid, and then standardized with respect to the 1961–2005 period statistics to reduce systematic biases in the mean and variance of AOGCM predictors relative to NCEP data. A drawback of the standardization approach is that it considers bias in the mean and variance only.

Table 4-1: CMIP5 Models used in the present study and their attributes

Modeling Center	**Institute ID**	**Model Name**
Canadian Centre for Climate Modelling and Analysis	CCCMA	CanESM2
Meteorological Office Hadley Centre	MOHC	HadGEM2-ES
Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	MIROC	MIROC-ESM
Max Planck Institute for Meteorology	MPI-M	MPI-ESM-LR
Norwegian Climate Centre	NCC	NorESM1-M
NOAA Geophysical Fluid Dynamics Laboratory	NOAA GFDL	GFDL-ESM2G

The following NCEP predictors were found by Asong et al. (2015) to influence significantly the historical climate characteristics both in space and time: 2-m air temperature, 850-hPa relative humidity, 500-hPa specific humidity, 850-hPa geo-potential height, mean sea level pressure, horizontal wind components (850-hPa meridional and 10-m zonal wind), and vertical velocity (i.e. omega at 500-hPa). In the current study, these predictors were extracted from AOGCM outputs and used alongside the calibrated GLMs for generating projected sequences of daily precipitation and temperature fields.

4.3 Methodology

This section provides a summary of the methodological background of the GLM framework for downscaling future projections of daily precipitation and temperature sequences. Most of the information provided in this section on GLMs is as implemented in the *Rglimclim* software package of Chandler (2014), which is used for this study.

4.3.1 GLM for Daily Precipitation

In a GLM, an $n \times 1$ vector of data $y_1, ..., y_n$ are considered to be the realized values of the random variables $\boldsymbol{Y} = (\boldsymbol{Y_1}, ..., \boldsymbol{Y_n})'$ with a mean vector $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)'$ where μ_i is related to the values of a row vector $\boldsymbol{x_i}$ of predictors such that:

$$g(\mu_i) = \boldsymbol{x_i}\boldsymbol{\beta} = \eta_i \tag{4-1}$$

where $g(.)$ is a monotonic transformation known as the link function and $\boldsymbol{\beta}$ is a $x \times 1$ vector of coefficients. The precipitation occurrence process (i.e., the pattern of wet and dry days) is modelled using logistic regression and the precipitation amount (i.e., intensity) on wet days is modelled using the gamma distribution. The precipitation occurrence process takes the form:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{x_i}\boldsymbol{\beta} \tag{4-2}$$

where p_i is the probability of precipitation for the i^{th} case in the dataset conditional on a covariate row vector $\boldsymbol{x_i}$ with coefficient column vector $\boldsymbol{\beta}$. Subsequently, for a potentially different covariate vector $\boldsymbol{\xi_i}$, the precipitation intensity process for the i^{th} wet day is modelled as gamma-distributed with mean μ_i and shape parameter ν for all observations at all sites (e.g., Yang et al., 2005), where

$$\ln(\mu_i) = \boldsymbol{\xi_i}\boldsymbol{\varphi} \tag{4-3}$$

$\boldsymbol{\varphi}$ is a column vector of coefficients. The coefficient vectors $\boldsymbol{\beta}$ and $\boldsymbol{\varphi}$ are estimated using the maximum likelihood method assuming that the observations from different sites are independent (Chandler and Bate, 2007), with subsequent adjustments for inter-site dependence.

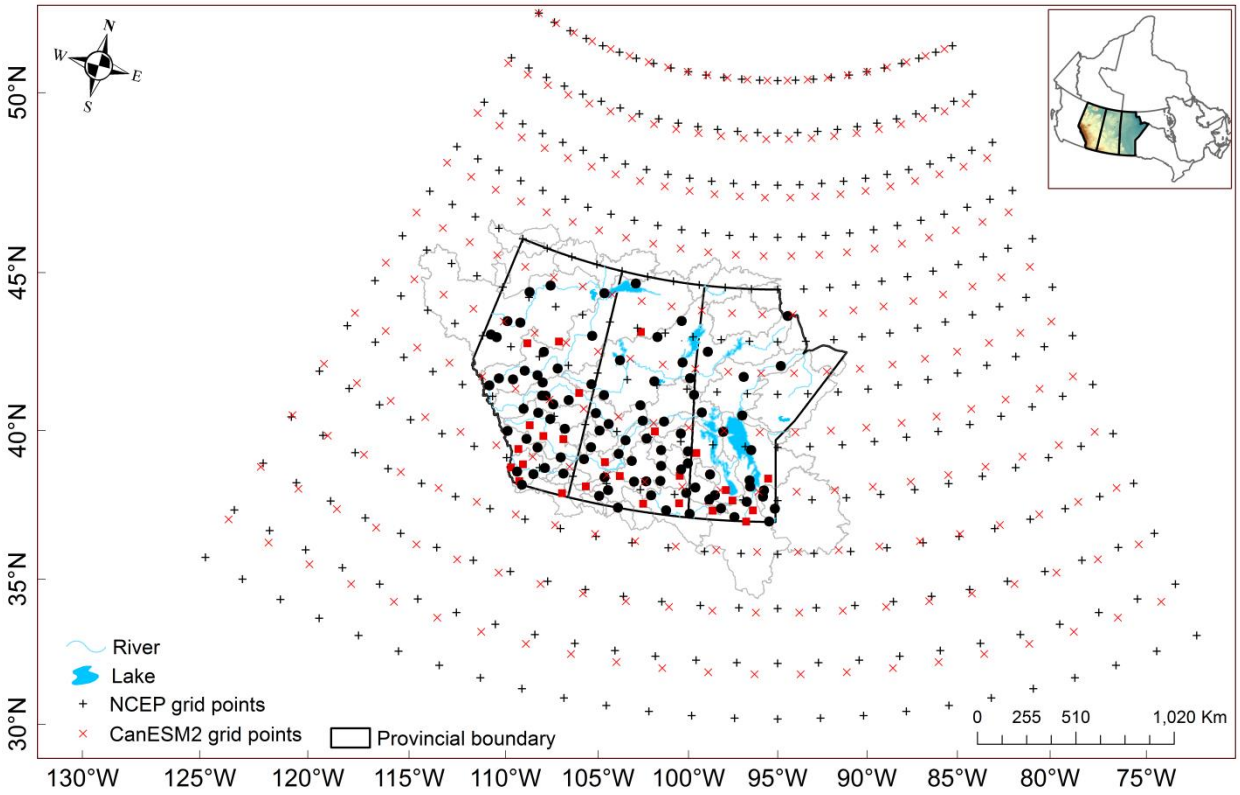

Figure 4-2: The spatial domain used for identification of predictor variables influencing weather processes in the Canadian Prairies Provinces. The NCEP variables are resolved at the 2.5^{o} x 2.5^{o} horizontal resolution (+ signs) while AOGCM variables were extracted over the same spatial domain. For example, the CanESM2 (2.8125° x 2.8125°) grid points (*x* signs) are indicated on the map alongside those of NCEP variables.

4.3.2 GLM for Daily Temperature

Several approaches have been used for modelling temperature sequences including a combination of a linear regression component to describe the linkage between predictors and temperature values, and a stochastic component based on a spatial moving-average process to reproduce the observed spatial dependence between the values at different sites (Khalili et al., 2013). Regression-based methods and artificial neural networks (Schoof and Pryor, 2001), first–

order trivariate auto-regression that is conditional on precipitation occurrence as implemented in Weather GENerator (WGEN) by Richardson and Wright (1984), and a first–order auto-regression as implemented by Chen et al. (2012) in the MulGETS WG have also been used. However, standard linear regression methods assume constant variance for daily time series, $\boldsymbol{Y}_{st}$, at each site s on a given day t. In most cases, the assumption of constant variance is often violated when analysing temperature series at the daily time scale (Chandler, 2014). Therefore, following Chandler (2005), the method used here includes a two-stage approach whereby separate mean and variance components are developed within a normal-heteroscedastic framework in which the mean (μ_{st}) and variance (σ^2) of $\boldsymbol{Y}_{st}$ depend on possibly different covariate vectors. For joint modelling of *Tmax* and *Tmin,* we model the mean of the two values using a normal distribution, and then the difference between them using a gamma distribution. This ensures that *Tmax* > *Tmin* always in any simulated sequences.

4.3.3 Training and Testing of the GLM Framework

In this study, GLMs were fitted using historical data (i.e., observed precipitation and temperatures as well as atmospheric predictors derived from NCEP reanalysis products). Then, for downscaling future projections under different scenarios, weather sequences were simulated from the fitted models driven by corresponding AOGCM atmospheric predictors. GLMs were fitted separately to precipitation and temperature observations considering the entire study area as a single region and using observations from the period 1971–2000. A wet day was defined as that with recorded amount of precipitation exceeding 0.5 mm. For the precipitation case, models were fitted using data from all 120 sites. Subsequently, *Tmin* and *Tmax* from 96 of the 120 stations were modelled separately and intervariable relationships were accounted for by using simultaneous and lagged values of precipitation as covariates to model temperature. The first

step involved in the downscaling model training was the development of 'initial' GLMs consisting of a constant term and basic factors influencing weather variability such as seasonality, autocorrelation and geographical attributes (site effects). Subsequently, daily values of NCEP predictors were incorporated as external covariates.

In order to simulate non-stationary weather sequences, some predictors were allowed to modulate the effect of others. This was incorporated via interactions (unlike fitting separate models for each month of the year). *Rglimclim* offers various options for modelling temporal autocorrelation structure mostly as a function of lagged values and a 'persistence indicator'. The seasonality structure in the data was modelled via various options (e.g., sine and cosine terms). A meaningful GLM for generating multisite multivariate weather sequences must preserve not only the temporal but also the spatial coherence. This requires a computationally tractable representation of inter-site dependence. This feature was incorporated by transforming the data values to Gaussianity and then studying inter-site correlations on the transformed scale (see Yang et al. (2005) for details).

Rglimclim provides a wide range of residual-based diagnostics to check that the fitted models are able to reproduce the systematic structure in the observations, as well as the distributional assumptions (e.g., precipitation intensities follow gamma distribution) and the assumed inter-site correlation structure. To check that the underlying structure has been captured by the fitted models, Pearson residuals were calculated as:

$$r_i^{(P)} = \frac{Y_i - \mu_i}{\sigma_i} \tag{4-4}$$

where Y_i is the observed response for case *i*, and μ_i and σ_i are the modelled mean and standard deviation. If the fitted model is reasonable, all of the Pearson residuals are expected to have mean zero and variance 1. In addition to Pearson residuals, Anscombe residuals (Equation 4-5)

for the gamma distribution are defined for the amounts model to ensure that the probability structure of the fitted models is correct.

$$r_i^{(A)} = \left(\frac{Y_i}{\mu_i} \right)^{1/3} \tag{4-5}$$

The suitability of the fitted models for generating weather sequences outside the training period was tested by using data for the pre- and post-fitting periods (i.e., 1961–1970 and 2001–2005). To do this, the parameters of the fitted models are constrained using NCEP predictors from the corresponding testing periods. Using the fitted models, 100 realizations were obtained for both training and testing periods. In each case, predictors for the first year were used to initialize the simulations. Subsequently, selected statistics such as the mean, standard deviation, lag-1 autocorrelation function, proportion of wet days, conditional mean and conditional standard deviation were computed from the simulated sequences and compared with the corresponding observed values. Conditional statistics were computed for precipitation only, based on the proportion of exceedances of the 0.5 mm threshold. It was important to assess the spatial distribution of precipitation and temperature as simulated by the GLMs for the 1961–2005 period versus observations. To this end, climatological mean daily values of precipitation and temperatures were computed for each site from 100 simulations and compared with observations.

4.3.4 Future Projection of Precipitation and Temperature Characteristics

The ultimate aim of this study was to simulate a range of future weather sequences that are consistent with future climates projected by the AOGCMs. Therefore, for each site, using the fitted model parameters, 100 simulations of daily precipitation and temperature sequences were made for the 1961–2005 historical period (conditioned on NCEP predictors), and for the 2006–2100 period conditioned on outputs from the six AOGCMS (Table 4-1) corresponding to

RCP2.6, RCP4.5, and RCP8.5 scenarios. However, for evaluating the future climate change signal at each site relative to the historical period, two non-overlapping 44-year time slices (i.e. 2011–2054 and 2055–2098) were considered. From the simulated distributions, the projected climate change signal was assessed for mean and extreme characteristics of precipitation and temperatures for all AOGCMs and RCP scenarios considered.

4.3.4.1 Mean characteristics

To evaluate projected changes in the mean climate between the historical and future periods, the climatological mean daily precipitation and temperatures were computed for each simulation at all sites. For an estimate of the relative change in a mean index from historical to future period, delta statistics were derived and converted to relative percentage changes. These changes were calculated for two 44-year future epochs: 2011–2054 (referred to hereafter as 2030s) and 2055–2098 (referred to hereafter as 2080s) relative to 1961–2005 (referred hereafter as 1980s). In order to test if the change from historical to future period is significant or not, the Student's *t-test* was applied. For significance testing, the 95% confidence level was used throughout the study. It has been re-emphasized that although significance testing aids in the interpretation of results, the significance of projected changes is difficult to interpret formally (e.g., Ambaum, 2010).

4.3.4.2 Extreme characteristics

For evaluating projected changes in extremes, some investigators have focused primarily on (1) various climate extremes indices which represent moderate extreme climatological events with re-occurrence times of a year or less (e.g., Tebaldi et al., 2006), and (2) asymptotic extreme value theory by employing the Generalized Extreme Value (GEV) distribution to approximate the distribution of seasonal extremes. This type of evaluation provides insights into the behaviour of extreme events with multi-year to multi-decadal return times that are of importance to, for

example, water resources systems design and planning. First, we assessed projected changes in commonly used climate indices, such as the mean wet spell length, mean dry spell length, consecutive dry days, consecutive wet days, extreme hot and cold temperature spells (i.e. the 90^{th} percentile heat wave duration and the 10^{th} percentile cold wave duration), number of frost days, and number of days without defrost (e.g., Goodess, 2003). Subsequently, we exploited asymptotic extreme value theory. Thus, seasonal maxima/minima was extracted for both *Tmax* (summer) and *Tmin* (winter) while 3-day extremes (i.e. heaviest three-consecutive day precipitation) were extracted for both winter and summer precipitation amounts for each simulation for each of the 120 sites. For example, during the 2030s, 3-day summer precipitation extremes were computed from each of the 100 simulations per AOGCM per site.

We employed the maximum likelihood (MLE) procedure for estimating the parameters of the GEV distribution that are functions of time. The location and scale parameters were assumed to depend linearly on time, while the shape parameter was assumed to be time-invariant. A return value for a specified *T*-year return period is the value that is exceeded by a seasonal extreme with probability $p = 1/T$. The analysis presented here was performed for the 20-year return period only which is chosen to be a compromise between the rareness of the event of interest and uncertainty associated with the estimated return values due to the limited sample size. The GEV distribution was fitted to each simulation separately, and then an average of the 20-year return values estimated at the end of each time window was obtained from all simulations, which in turn was used in estimating the relative change. The uncertainty (in the form of standard error) associated with the averaged return value was obtained using an error analysis approach from Bevington and Robinson (2002). The standard error thus obtained was used to develop the 95% confidence interval under the normality assumption. In a similar manner, based on the concept of error analysis, the standard errors and associated confidence intervals were derived for multi-model

ensemble averaged return values. Finally, for a given site, if the confidence intervals of the 20-year return level estimates between the historical and future periods do not overlap, then the change was judged as statistically significant.

4.4 Results and Discussion

The results of various analyses based on the approaches described in the above section are presented here. An evaluation of the performance of the GLMs in terms of reproducing spatial structure of mean daily precipitation and minimum and maximum temperatures is presented first followed by an assessment of projected changes to mean and extreme characteristics of both precipitation and temperature fields relative to the historical period simulations generated using the trained models. Detailed evaluation of GLMs is presented in Asong et al. (2015, under review) (i.e. Chapter 3 of this thesis).

4.4.1 Evaluating GLM-Simulated Mean Climate during the Historical Period (1962 – 2005)

Figure 4-3 shows the spatial structure of observed and simulated mean daily precipitation and their differences for the 1962–2005 historical period. The modelling approach does reasonably well in reproducing the observed spatial patterns of precipitation. For all seasons, the GLMs tend to slightly overestimate precipitation values in the middle to northern watersheds while underestimate it in the western to eastern watersheds. Overall, the simulations indicate a dry bias in winter (-0.002 mm), spring (-0.05 mm), summer (-0.2 mm), and autumn (-0.1 mm) as averaged over all 120 sites. The simulated and observed *Tmin* is shown in Figure 4-4. There is good correspondence between the modelled and observed *Tmin* values with typical discrepancies lying within about -0.3 ^{o}C and +1.5 ^{o}C. On a seasonal basis, the spatial evolution of *Tmin* is well captured in spring relative to the other seasons. A warm bias (overestimation) is evident in winter (1.1 ^{o}C), spring (0.2 ^{o}C) and summer (0.4 ^{o}C) while a cold bias (underestimation) exists in autumn (-0.1 ^{o}C). The results for *Tmax* are shown in Figure 4-5. Overall, the GLMs reproduced

satisfactorily the spatial distribution of observed *Tmax* values. However, the models over (under) simulated temperature values in winter (0.6 ^{o}C) and spring (0.5 ^{o}C) (summer: -1.2 ^{o}C and autumn: -0.5 ^{o}C). This performance is judged to be satisfactory given that temperature is simulated at all sites including 24 ungauged sites and 96 gauged sites used in fitting the models. This alone, in reality should have increased the margin of error. These results cannot be validated against other multisite multivariate modelling studies given that to the best of our knowledge, so far, this study is the first of its kind in this region of Canada. The biases reported here are within reasonable limits of any plausible multisite multivariate stochastic modelling framework applied at such a large spatial scale (e.g., Wilks, 1998).

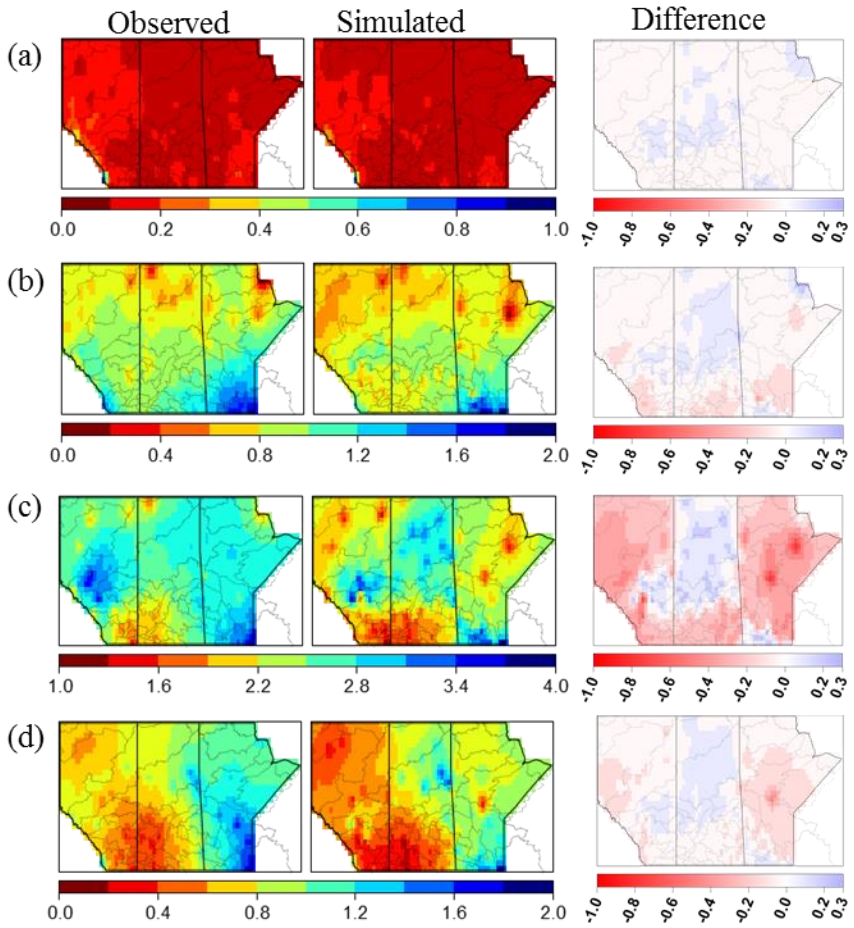

Figure 4-3: The spatial structure of observed (left) and simulated (middle) daily mean precipitation climatology (mm/day) and their difference (right) for (a) winter (December–February), (b) spring (March–May), (c) summer (June–August), and (d) autumn (September–November) seasons over the 1962–2005 period.

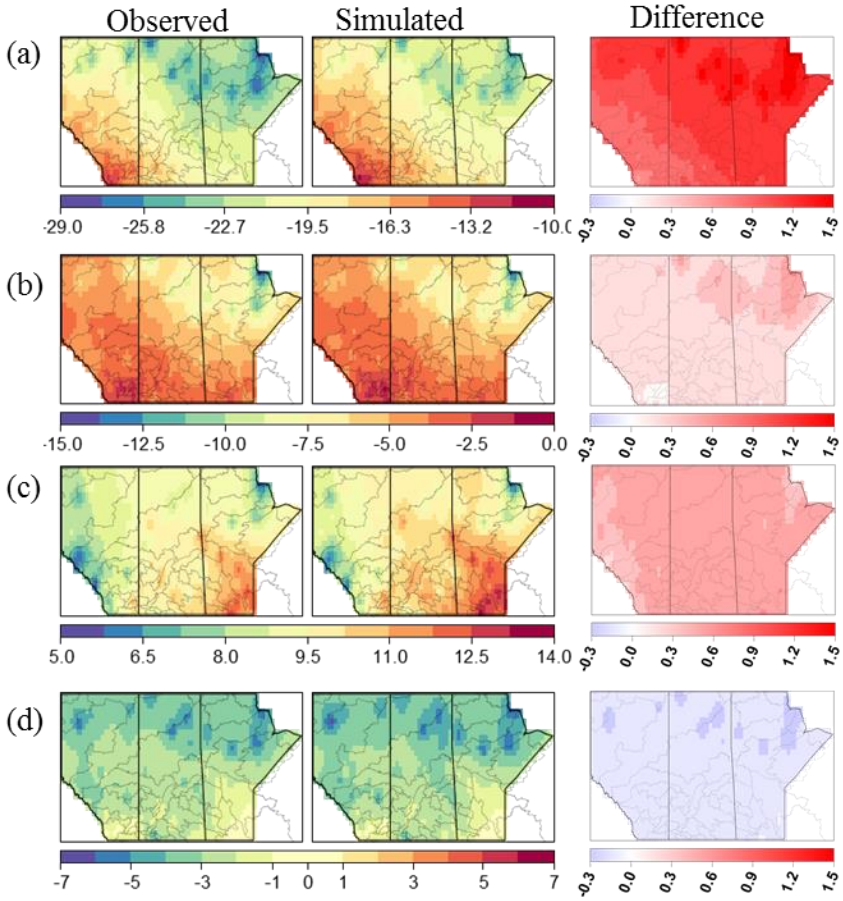

Figure 4-4: Same as in Figure 4-3 but for mean daily minimum temperature (°C/day) for the 1962–2005 period.

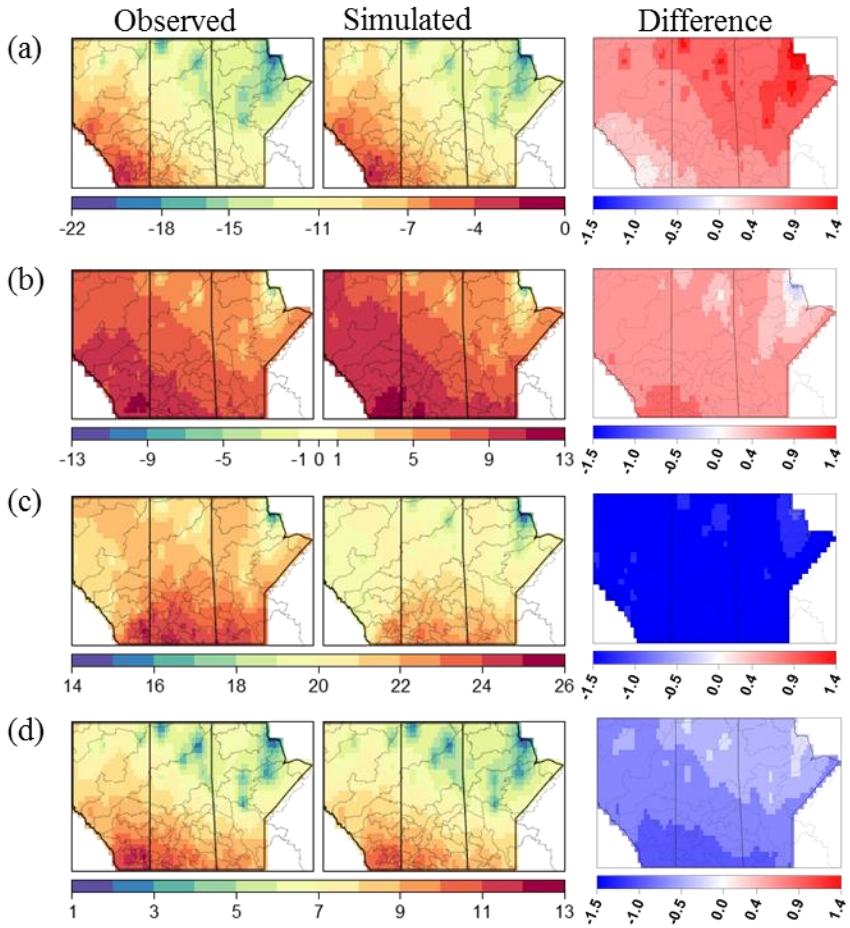

Figure 4-5: Same as in Figure 4-3 but for mean daily maximum temperature (°C/day) for the 1962–2005 period.

4.4.2 Projected Changes in Mean Characteristics

4.4.2.1 Temporal evolution

The analysis starts with a discussion of the temporal evolution of downscaled seasonal daily precipitation and temperatures as pooled over all sites in the study region. Figure 4-6a shows simulated percentiles of downscaled daily precipitation in summer (JJA) shown for all AOGCMs used in the study. Relative to the 1962–2005 reference period, all AOGCMs as well as the multi-model ensemble (MME) project a general increase in precipitation in the 21st century. However, HadGEM2-ES projects a likely decrease in summer precipitation for the period 2060–2080. The MME median increase in precipitation over the 120 sites projected by the end of the

21st century is 7%, 5.8% and 4% for RCP2.6, RCP4.5 and RCP8.5, respectively. Although there is an overall increase in precipitation in the summer, the response is weaker with increased GHG concentration. By comparing the future and historical simulations, the range of the simulated precipitation is wider for the future than the historical period implying that future projections may be more uncertain, suggesting that individual AOGCM projections may be divergent (i.e., all AOGCMs do not totally agree). RCP8.5 seems to have a larger associated variability, indicating that the sensitivity of future projections on average increases with the GHG concentration. Unlike summer, most AOGCM projections indicate an overall slight decrease in winter precipitation (Figure 4-6b) with significant disagreement in terms of the range of values among different AOGCMs and RCPs. The MME median decrease in precipitation projected by the end of the 21st century is 1.8%, 2.3% and 3.1% for RCP2.6, RCP4.5 and RCP8.5, respectively.

Concerning the temporal evolution in temperatures, all AOGCMs project a likely increase in summer *Tmax* (Figure 4-7a) and winter *Tmin* (Figure 4-7b) under all RCPs by the end of the 21st century. The projections indicate that *Tmin* is more variable than *Tmax*, and therefore future projections relative to the historical period seem to be more uncertain. The MME median increase in *Tmax* (*Tmin*) projected by the end of the 21st century is 9.3 (10), 11.7 (12.4) and 14 (16.2) % for RCP2.6, RCP4.5 and RCP8.5, respectively. Unlike precipitation, associated variability in temperatures is relatively small with increased radiative forcing as derived using the MME (last row in Figures 4-6a, b and 4-7a, b). We further examined the evolution of precipitation and temperatures in spring and autumn (Figure 4-8) as projected using the MME. Although with increased variability, the projections indicate a tendency towards an increase in precipitation, *Tmin* and *Tmax* by the end of the 21st century under all RCP scenarios. For example, relative to the historical period, the projected increase in precipitation, *Tmin* and *Tmax*

by the end of the 21st century in spring (autumn) is 3.2 (2.8), 4.8 (4.3) and 4.4 (4.1) % under RCP8.5. Just like for other seasons, the increase is greater with increased concentrations of GHGs in the atmosphere.

Generally, most statistical downscaling models are trained and tested by using reanalysis products (e.g., NCEP) and observations corresponding to the historical climate. Then, for generating future projections, outputs of a certain AOGCM pertaining to a selected RCP scenario are introduced to the downscaling model. This procedure in most cases does not provide a smooth transition from the downscaling model development phase to the future projection phase, as the simulations for the different phases are performed with the outputs of two different sources. Thus, the inputs used in the development phase and the future projection phase of the downscaling models are not homogeneous. The PP SD approach used here assumes that the predictors used for training the SD model are also realistically simulated by the target AOGCM. Although our temporal plots satisfy this assumption, there is a difference in the variability of the two time series. In Figure 4-8, the change in the width of the distribution in the time series plots at the transition from the past to the future period might be due to the fact that for the past one realization of predictors is used, whereas for the future it is a MME.

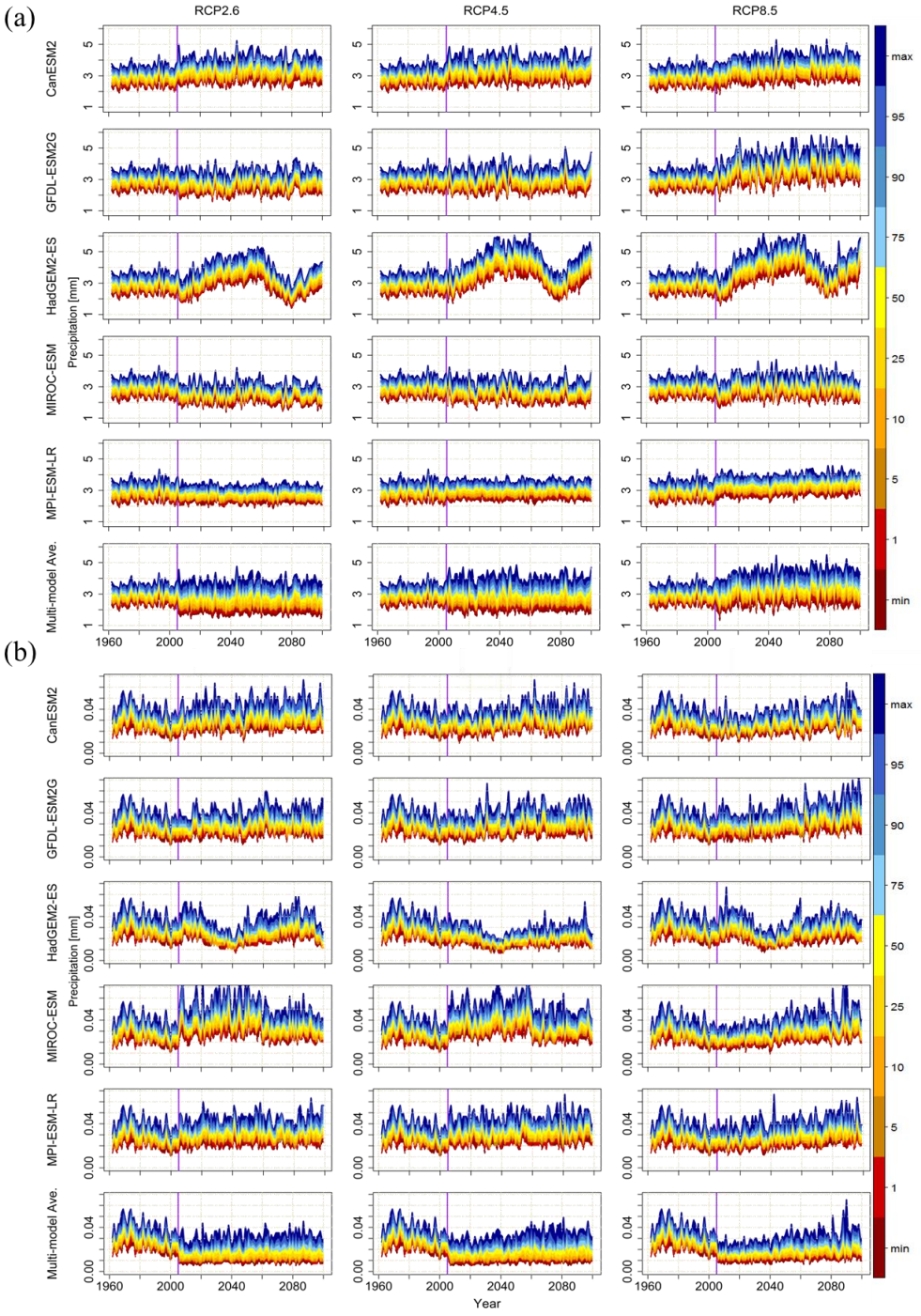

Figure 4-6: Temporal evolution of downscaled daily mean precipitation in summer (a) and winter (b) for the historical (1962–2005) and future (2006–2098) periods, pooled over all sites in the study area, for five of the six AOGCMs considered in the study. The last row in each season corresponds to multi-AOGCM ensemble average, while columns correspond to three RCP scenarios. The purple line in each panel splits the historical and future periods. Colour bands (from red to dark blue) indicate minimum and maximum values together with the 1st, 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles of simulated precipitation amounts.

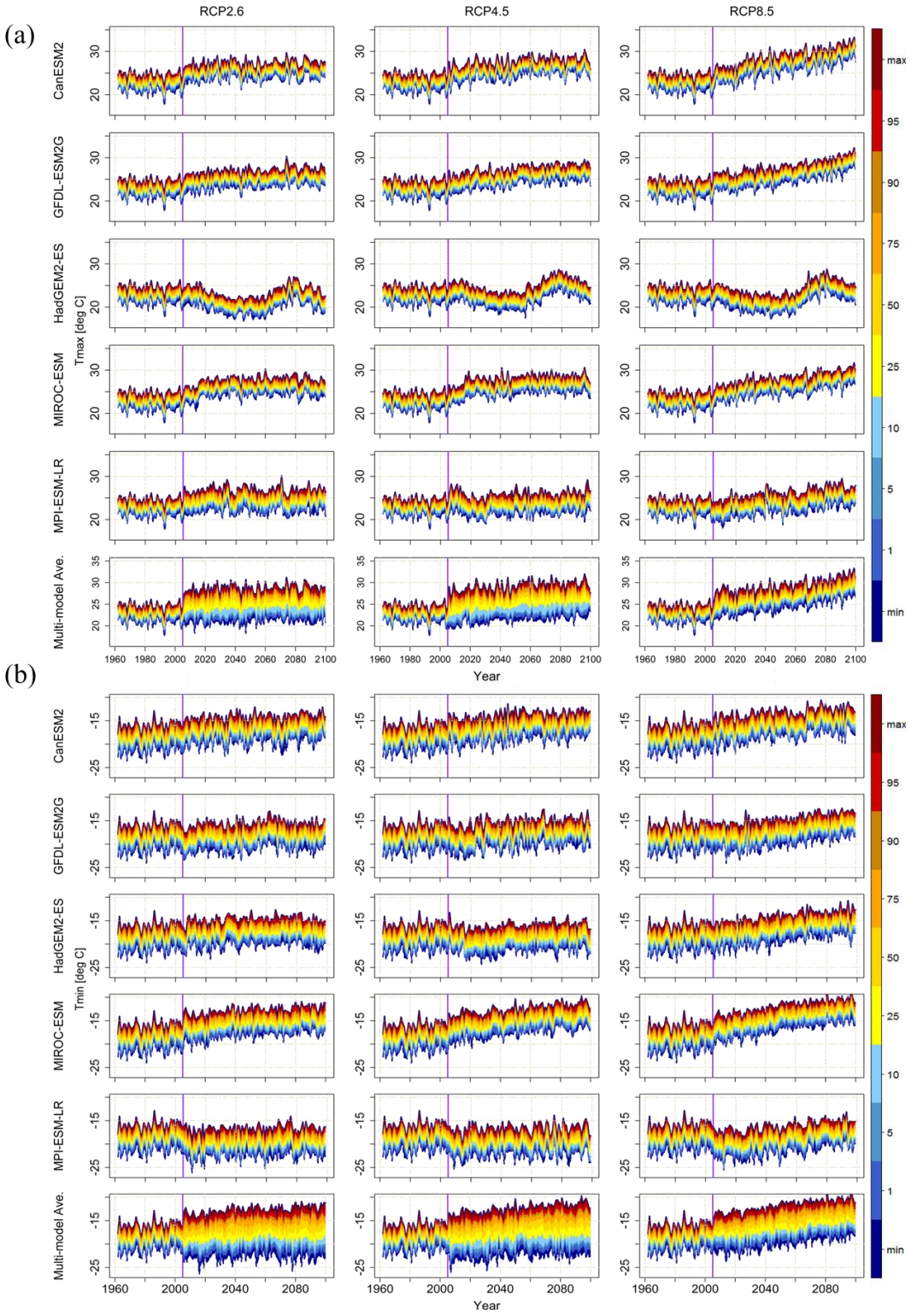

Figure 4-7: Same as in Figure 4-6 but for the mean daily maximum temperature for the summer (a) and mean daily minimum temperature in the winter (b) season.

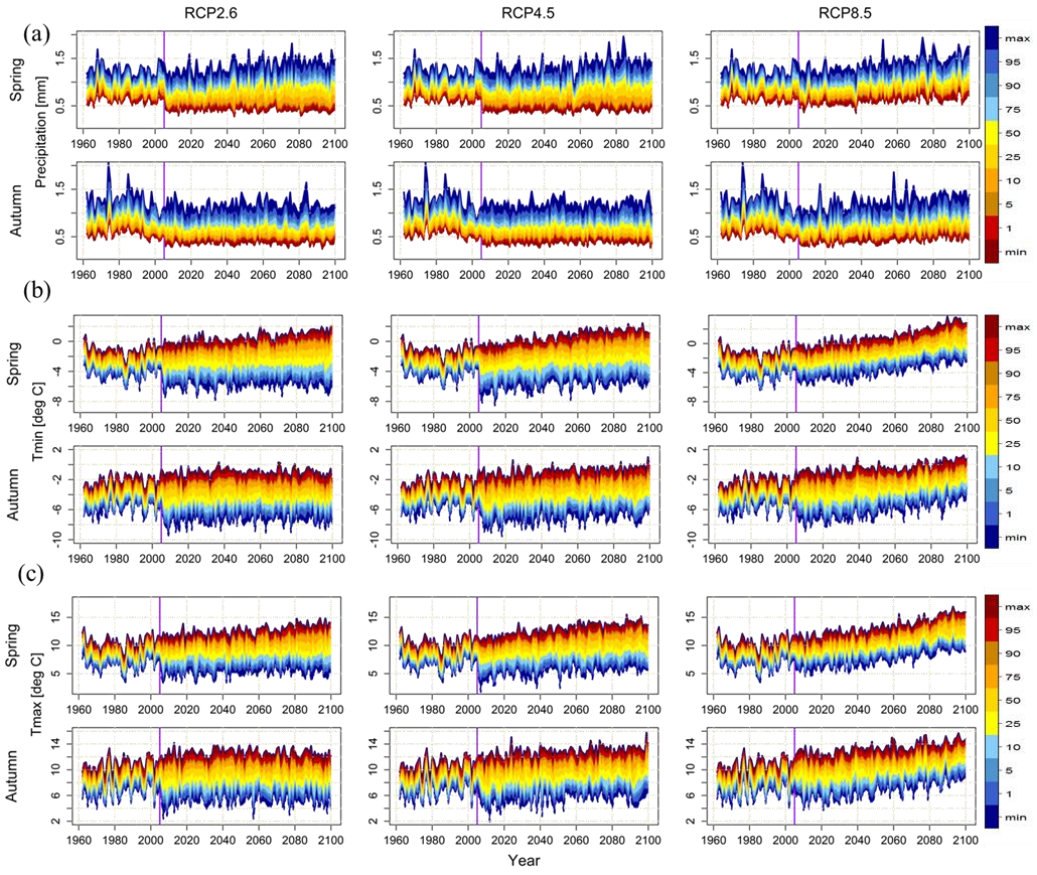

Figure 4-8: Multi-model ensemble averaged (a) mean daily precipitation, (b) *Tmin*, and (c) *Tmax* in spring and autumn. All other information is the same as in Figure 4-6.

4.4.2.2 Spatial and seasonal patterns

To understand projected dynamics of precipitation and temperature regime changes across the study area spatially, mean climate is computed for the future periods and compared with the historical climate. Results in terms of percentage changes are shown for both winter and summer for the MME of AOGCM-driven GLM simulations as well as for individual AOGCM simulations. Figure 4-9 depicts projected changes in mean daily precipitation for the 2030s and 2080s relative to the 1980s. In winter (Figure 4-9a and Table 4-2), precipitation is projected to decrease over the entire region for the three RCPs considered for the 2030s and 2080s. The decrease is greater in the 2080s than the 2030s and more so with the magnitude of GHG concentration. For instance, the MME areal-averaged change in precipitation in the 2030s

(2080s) is -4 (-4.8), -3 (-12), and -10 (-23) %, respectively under RCP2.6, RCP4.5 and RCP8.5 scenarios, respectively. In Table 4-2, although with different sensitivities to the choice of RCP, all AOGCMs project a probable decrease in precipitation in winter and the change signal is greater with GHG concentrations. However, only changes in the 2080s under RCP8.5 are found to be significant at a critical p-value of 0.05. Unlike in winter, the projections show a significant increase in precipitation in summer (Figure 4-9b and Table4-2) under all RCP scenarios. The projections indicate that the 2030s will probably be wetter on average than the 2080s and less so with increased GHG forcing. The MME projected increase in mean precipitation over all sites in the 2030s (2080s) is 10.3 (9.8), 9 (8.2), and 9.4 (7) %, respectively for RCP2.6, RCP4.5 and RCP8.5 scenarios. Concerning regional patterns of change, the least increases are expected to occur over southern regions pertaining to the Lake Winnipeg, Lake Manitoba, Athabasca River, and North Saskatchewan River watersheds. This signal is consistent over all scenarios and temporal periods. Similar patterns are found when AOGCMs are considered individually, although the range of changes varies substantially from one model to the other (Table 4-2) except CanESM2 which seemed insensitive to the choice of RCP.

The MME projected changes in *Tmax* are shown in Figure 4-9b and Table 4-3 for individual AOGCMs for both winter and summer seasons. Generally, average maximum daily temperature is projected to increase in both seasons over the entire study area for the 2030s and 2080s. The projected change is most likely to intensify in the 2080s than the 2030s and more so for winter than for summer especially for RCP8.5. The most warming is simulated in the high emissions scenario (RCP8.5). In winter (summer), for the 2030s, the areal-averaged percentage change in *Tmax* is 5 (10), 9 (12), and 14 (12.6) % while for the 2080s, this change is 15 (9), 13.8 (13), and 26 (22.5) % under the RCP2.6, RCP4.5 and RCP8.5 scenarios, respectively. In Table 4-3, CanESM2 tend to be more sensitive to the choice of RCP unlike MIROC-ESM and other

AOGCMs. Regionally, the strongest warming in *Tmax* generally occurs in the north-eastern watersheds, such as in Nelson River, Seal River, Kazam River and Hayes River.

Table 4-2: Projected changes (%) in the mean daily precipitation as downscaled from individual AOGCM outputs under three RCP scenarios for the 2011–2054 and 2055–2098 periods relative to the 1962–2005 reference historical period. The range of change across the 120 sites considered in this study is indicated in brackets while the sign in front of the brackets shows the direction of change.

Winter	**2011–2054**			**2055–2098**		
	RCP2.6	**RCP4.5**	**RCP8.5**	**RCP2.6**	**RCP4.5**	**RCP8.5**
CanESM2	- (1–5)	- (2–3)	- (2–15)	- (3–7)	- (2–8)	- (2–18)
HadGEM2-ES	- (1–9)	- (4–10)	- (4–21)	- (2–8)	- (4–11)	- (2–28)
MIROC-ESM	- (1–6)	- (2–8)	- (2–14)	- (3–8)	- (5–11)	- (2–19)
MPI-ESM-LR	- (2–5)	- (1–6)	- (4–16)	- (3–7)	- (2–16)	- (3–25)
NorESM1-M	- (1–6)	- (2–6)	- (2–11)	- (1–8)	- (2–9)	- (2–28)
GFDL-ESM2G	- (2–5)	- (2–7)	- (2–18)	- (3–6)	- (2–8)	- (5–26)
Summer						
CanESM2	(9–15)	(8–16)	(6–17)	(11–16)	(10–17)	(10–18)
HadGEM2-ES	(8–12)	(9–11)	(9–15)	(9–13)	(10–16)	(10–19)
MIROC-ESM	(7–9)	(7–11)	(8–15)	(8–12)	(9–14)	(11–22)
MPI-ESM-LR	(7–11)	(6–10)	(10– 15)	(7–11)	(8–17)	(9–17)
NorESM1-M	(11–14)	(12–16)	(8–17)	(12–16)	(11–19)	(7–23)
GFDL-ESM2G	(8–13)	(7–14)	(8–17)	(7–13)	(8–15)	(9–17)

Table 4-3: Projected changes (%) in the mean daily *Tmax* as downscaled from individual AOGCM outputs under three RCP scenarios for the 2011–2054 and 2055–2098 periods relative to the 1962–2005 reference historical period. Other information is the same as in Table 4-2

Winter	**2011–2054**			**2055–2098**		
	RCP2.6	**RCP4.5**	**RCP8.5**	**RCP2.6**	**RCP4.5**	**RCP8.5**
CanESM2	(1–9)	(1–10)	(1–13)	(2–11)	(6–24)	(5–38)
HadGEM2-ES	(2–11)	(2–11)	(2–16)	(1–13)	(5–23)	(6–36)
MIROC-ESM	(1–12)	(3–14)	(5–18)	(8–13)	(9–14)	(11–22)
MPI-ESM-LR	(7–11)	(6–10)	(10– 15)	(7–14)	(8–25)	(19–36)
NorESM1-M	(1–15)	(4–16)	(9–21)	(12–17)	(9–30)	(23–39)
GFDL-ESM2G	(3–13)	(4–14)	(7–23)	(5–15)	(7–19)	(17–37)
Summer						
CanESM2	(1–10)	(1–11)	(1–13)	(2–12)	(6–25)	(5–37)
HadGEM2-ES	(3–11)	(4–12)	(4–18)	(2–13)	(5–23)	(5–34)
MIROC-ESM	(3–13)	(4–15)	(5–19)	(6–15)	(8–16)	(10–21)
MPI-ESM-LR	(6–10)	(7–11)	(9– 14)	(6–14)	(9–15)	(18–34)
NorESM1-M	(9–14)	(9–17)	(10–22)	(13–14)	(14–27)	(21–33)
GFDL-ESM2G	(4–15)	(7–15)	(9–24)	(10–16)	(10–17)	(16–36)

The results of projections for average daily minimum temperature are shown in Figure 4-9c as well (and Table 4-4 for individual AOGCMs). The spatial patterns of changes in *Tmin* in most cases are similar to those for *Tmax* but the magnitude of change is different. In winter (summer), for the 2030s, the areal-averaged percentage change in *Tmin* is 7.3 (3.5), 8 (5.2), and 13 (6) % while for the 2080s, this change is 18.8 (6), 19.6 (7), and 37 (14.5) % under the RCP2.6, RCP4.5 and RCP8.5 scenarios, respectively. In particular, changes in winter *Tmin* are stronger than *Tmax* while *Tmax* will probably warm faster than *Tmin* in summer. In Figure 4-9, the greatest warming in *Tmin* exceeding 40% occurs in winter for the 2080s under RCP8.5 in such regions as the rocky Mountains, the northern parts of the Saskatchewan River basin, most northern watersheds and the areas to the southeast of Manitoba such as Winnipeg River watershed. However, in the summer, stronger warming exceeding 20% relative to the historical climate is anticipated for the 2080s under RCP8.5 for the Peace and Athabasca watershed regions, and for the middle to southern parts of the Saskatchewan River basin. The changes in mean temperatures are found to be statistically significant. In Table 4-4, the projected changes in *Tmin* differ considerably across individual AOGCMs. This finding is in line with Chapter 9 of the IPCC Assessment Report 5 (AR5) which also reports large inter-model spreads for the CMIP5 AOGCMs.

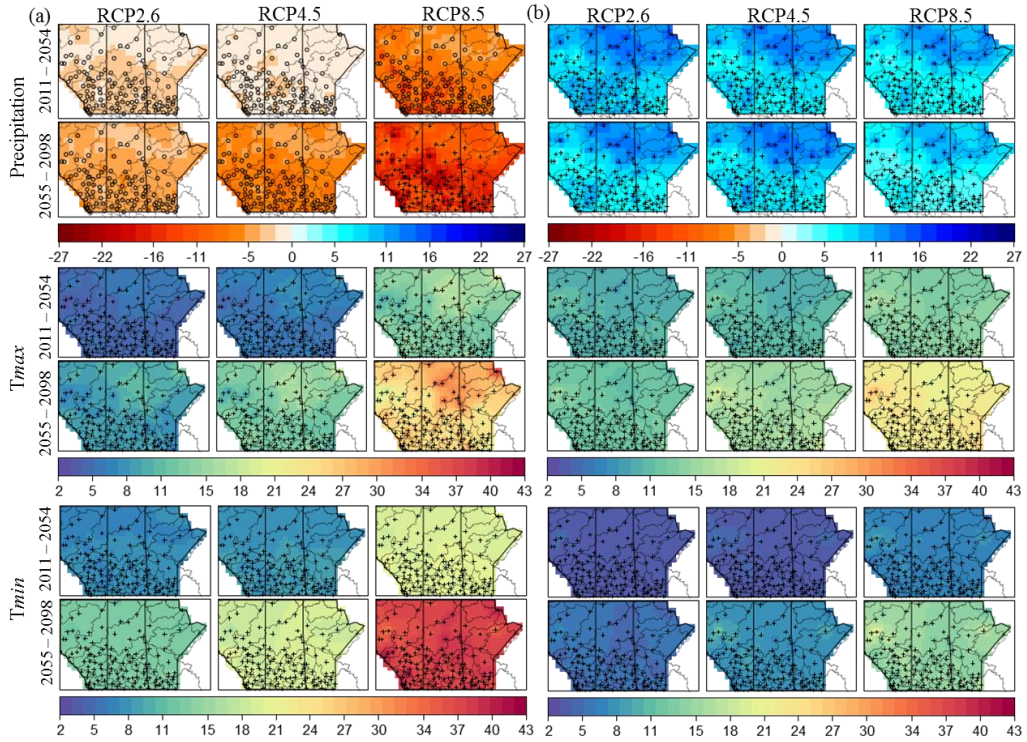

Figure 4-9: Multi-model averaged projected changes (in %) in daily mean precipitation, T*min* and T*max* for (a) winter and (b) summer for the 2011–2054 and 2055–2098 future periods with respect to the historical 1962–2005 period. Results are shown for the mitigation (RCP2.6), stabilization (RCP4.5) and high emissions (RCP8.5) scenarios. Pluses (dots) indicate stations where changes are found statistically significant (insignificant) at the 5% significance level.

Table 4-4: Projected changes (%) in the mean daily *Tmin* as downscaled from individual AOGCM outputs under three RCP scenarios for the 2011–2054 and 2055–2098 periods relative to the 1962–2005 reference historical period. Other information is the same as in Table 4-2

Winter	**2011–2054**			**2055–2098**		
	RCP2.6	**RCP4.5**	**RCP8.5**	**RCP2.6**	**RCP4.5**	**RCP8.5**
CanESM2	(1–7)	(1–10)	(2–11)	(2–14)	(6–23)	(17–44)
HadGEM2-ES	(3–6)	(4–13)	(2–15)	(3–9)	(2–29)	(7–40)
MIROC-ESM	(4–7)	(5–14)	(6–26)	(2–7)	(7–20)	(15–46)
MPI-ESM-LR	(2–11)	(5–11)	(7– 19)	(3–13)	(8–36)	(13–41)
NorESM1-M	(3–12)	(4–20)	(10–29)	(3–15)	(8–27)	(12–43)
GFDL-ESM2G	(1–10)	(2–16)	(4–27)	(2–16)	(6–20)	(14–42)
Summer						
CanESM2	(2–6)	(1–7)	(4–9)	(2–9)	(3–12)	(4–17)
HadGEM2-ES	(4–10)	(4–13)	(5–16)	(4–13)	(6–16)	(5–21)
MIROC-ESM	(1–11)	(3–12)	(3–18)	(3–19)	(4–18)	(5–19)
MPI-ESM-LR	(3–10)	(5–12)	(5– 13)	(4–14)	(6–14)	(3–18)
NorESM1-M	(2–12)	(4–16)	(5–14)	(4–12)	(6–21)	(4–22)
GFDL-ESM2G	(1–7)	(4–9)	(5–14)	(6–11)	(4–15)	(6–23)

4.4.3 Projected Changes in Climate Extremes

4.4.3.1 Climate indices

A comparison of changes in selected climate indices is presented here. The distributions of MME changes pooled over all sites in the study area are shown in Figure 4-10, both for precipitation (a) and temperature (b) indices evaluated on an annual basis. Changes are displayed separately for each period (i.e., 2011–2054 and 2055–2098) and for each of the three GHG concentration scenarios (i.e., RCP2.6, RCP4.5 and RCP8.5). For precipitation-based indices, conservative wet days (*pxcwd*) and mean wet spell lengths (*pwsav*) are projected to increase in the future for all scenarios. The strongest median increase in *pxcwd* of about 15% occurs in RCP8.5 likewise that for *pwsav* of about 6.5% in the 2080s. Consequently, dry days (-7%) and dry spell lengths (-24%) will likely decrease during the 2080s. In comparison, boxplots for the 2080s are more dispersed than for the 2030s. For example, the interquartile range of mean dry spell length in the 2030s under RCP8.5 is about -17% to -22% while it is approximately -23% to -30% in the 2080s. This indicates that 2080s projections may be more variable and uncertain. The response for RCP8.5 scenario appears to have the largest associated variability. Thus, the variability of future projections will likely increase with the magnitude of the GHG forcing.

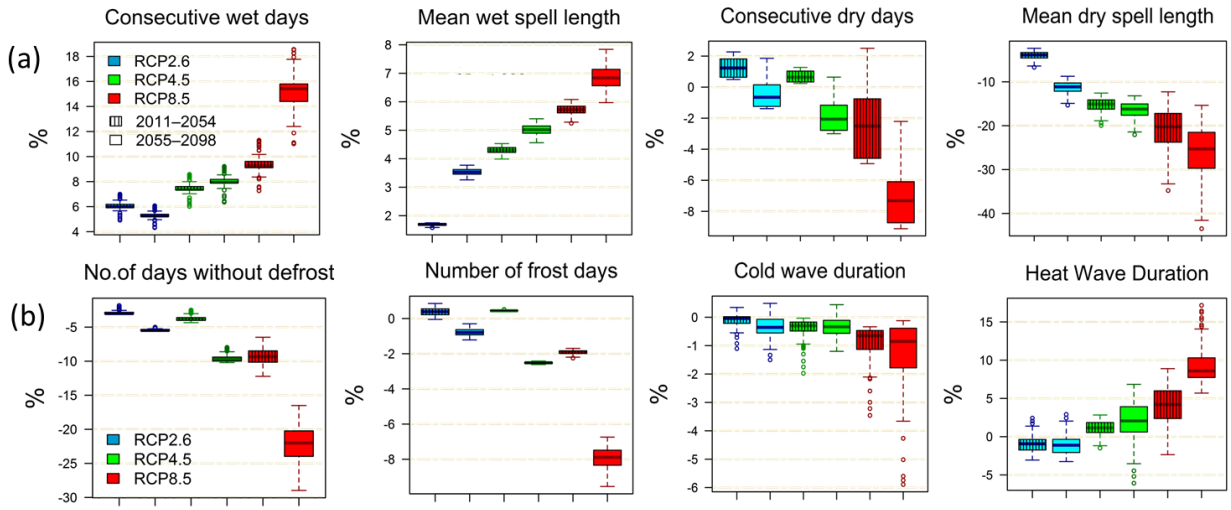

Figure 4-10: Multi-model projected changes (in %) in areal averaged (a) annual precipitation and (b) temperature indices over the periods 2011–2054 and 2055–2098 under RCP2.6 (blue), RCP4.5 (green), and RCP8.5 (red) relative to the historical period (1962–2005). Boxes indicate the interquartile spread (25^{th} and 75^{th} quantiles) while the black horizontal line indicates the median and the whiskers show the range of change.

Figure 4-10b depicts future projections of temperature-based indices. The projected increase in average daily temperature will likely be followed by a decrease in the proportion of frost days: Tmin < 0 (median decrease of -8% for RCP8.5) and days without defrost: *Tmax* < 0 (median decrease of -22% for RCP8.5) especially for the RCP4.5 and RCP8.5 scenarios. This indicates that it will likely snow less in winter in the future, thus, a decrease in snow-dominated daily precipitation during this season. Likewise, in terms of hot (the 90^{th} percentile heat wave duration) and cold temperature spells (the 10^{th} percentile cold wave duration), hot spells are projected to likely increase while cold spells will decrease with increased radiative forcing. The 2080s response is greater than the 2030s and the heat wave duration index under RCP8.5 tends to increase (median change of 7%) more than the cold wave duration index (~-1%) in the 2080s. Overall, the greater the radiative forcing, the larger the associated variability (spread). A higher likelihood of heat waves for instance may cause health related issues for populations at risks while a decrease in winter snowfall and increase in temperatures is likely, among other factors, to impact river flows, agriculture and other socio-economic and environmental sectors.

4.4.3.2 Frequency Analysis

An important feature of climatic variables is the behaviour of the tail of the distribution (i.e., extreme events). For a given site, return levels for each of the 100 GLM simulations and for each AOGCM for the historical and future periods are estimated by fitting the GEV distribution to the extracted extreme values, as explained in section 4.3.4.2. The projected changes in the 20-year return levels of 3-day seasonal precipitation extremes for the 2030s and 2080s relative to the 1980s are displayed in Figure 4-11 for RCP2.6, RCP4.5 and RCP8.5 scenarios. The results are

displayed for five of the six AOGCMs as well as for the MME. Relative to the historical period, winter season (Figure 4-11a) 3-day return levels will decrease in the future by the end of the 21st century, more so for the 2080s than the 2030s and that the magnitude of decrease appears to be greater with increased radiative forcing values. The areal-averaged MME projected decrease in 20-year return levels over all sites in the 2030s (2080s) is 4.3 (4.8), 5.4 (6.2), and 7.4 (-28) %, respectively for RCP2.6, RCP4.5 and RCP8.5 scenarios. The response is more pronounced for the 2080s compared to the 2030s. Substantial inter-model variability is noticeable. For example, HadGEM2-ES projects larger decreases in winter precipitation extremes than the other models. Spatially, these snow-dominated extremes will probably decrease more over the middle to southern parts of the study area where changes exceeding -26% are projected for RCP8.5 compared to northern parts. More importantly, changes in the 20-year return values of extremes largely exceed changes in mean precipitation in winter. Using RCP8.5 as an example, for the 2080s, the projected MME areal-averaged decrease in winter mean precipitation is about -23% compared to -28% in the case of extremes.

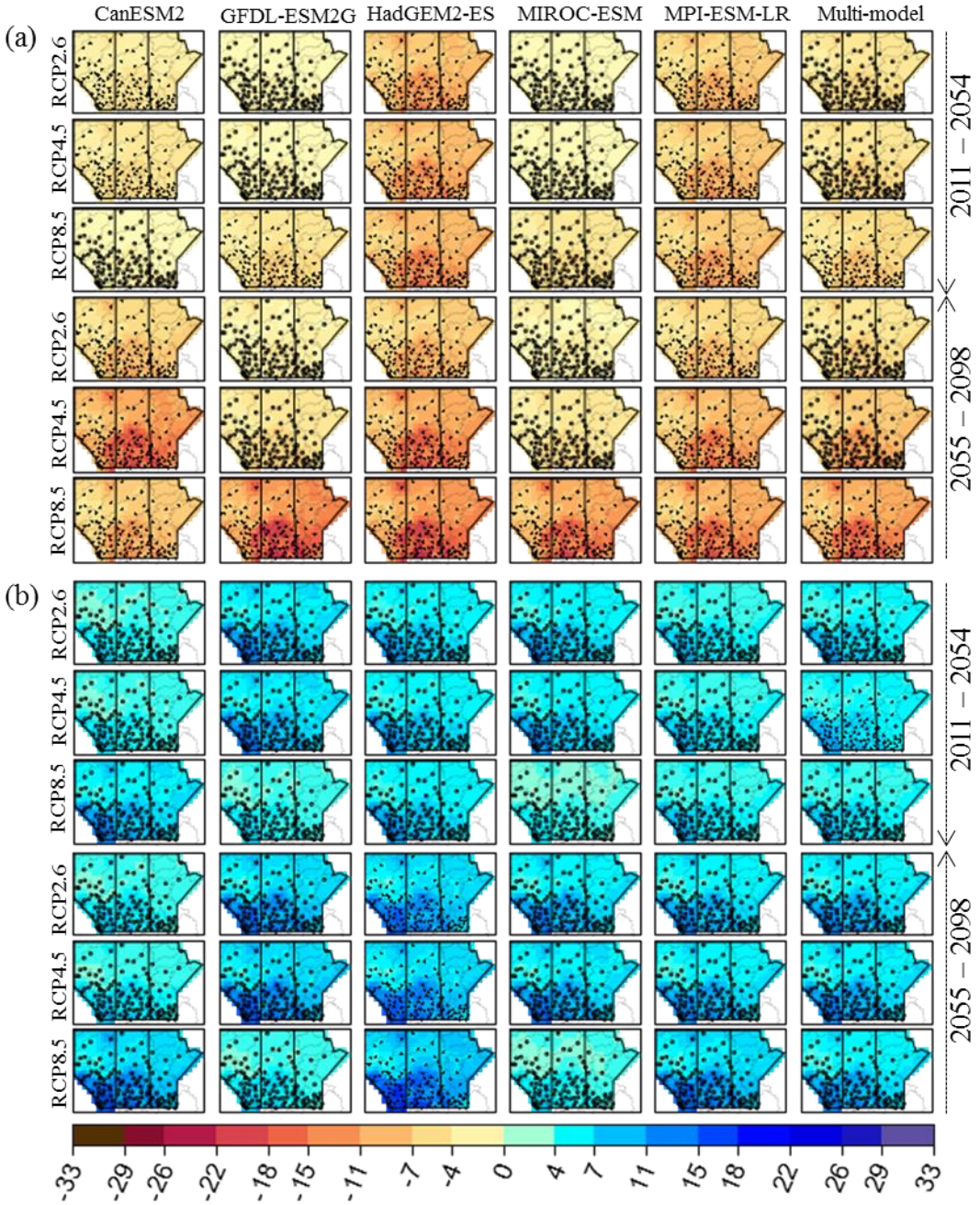

Figure 4-11: Individual and multi-model projected change (%) in 20-year return levels of 3-day precipitation extremes in (a) winter and (b) summer for the 2011–2054 and 2055–2098 periods for the RCP2.6, RCP4.5, and RCP8.5 experiments relative to the historical period (1962–2005). All other information is the same as in Figure 4-9.

Unlike winter extremes, the downscaled AOGCM projections point towards an increase (2–17%) in summer precipitation extremes (Figure 4-11b) across the study region under all RCP scenarios. Again, for the 2030s (2080s), the MME projected changes in the 20-year return values of extremes 11.7 (15%) exceed those in the mean precipitation 9.4 (7%) under RCP8.5. Based on the data used in this study, it is observed that the higher the forcing is, the higher the changes in return levels are although significant differences exist across individual AOGCMs. This is a mitigation scenario whereby radiative forcing will first increase sharply and then start to decrease by the 2050s (Moss et al., 2010). Generally, the higher the forcing is, the higher the changes in return levels are although significant differences exist across individual models.

Comparison of cold and warm temperature extremes predicted from the AOGCM-driven GLM simulations for the historical and future periods is shown in Figure 4-12. Overall, warm extremes (Figure 4-12b) are projected to intensify with global warming and cold extremes (Figure 4-12a) will likely warm faster than the warm extremes by the end of the 21st century under RCP8.5 scenario. Relative to the 1980s, 20-year return values of temperature extremes will likely increase over all sites and RCPs in the range of about 0–50% and appear to intensify with the magnitude of radiative forcing. The areal-averaged MME projected increase in 20-year *Tmin* return levels in the 2030s (2080s) is 18 (18.8), 21 (23), and 29 (34) %, respectively for RCP2.6, RCP4.5 and RCP8.5 scenarios. While the areal-averaged MME projected increase in 20-year *Tmax* quantiles in the 2030s (2080s) is 15 (17), 17 (21), and 24 (30) %, respectively for RCP2.6, RCP4.5 and RCP8.5 scenarios. Generally, the direction of the projected change signal is consistent among individual climate models but the magnitude of change is particularly different. For example, in winter (Figure 4-12a), for the 2030s, CanESM2 projects larger increases under RCP2.6 and RCP4.5 than RCP8.5. Thus, a multi-model approach is warranted in order to reduce the uncertainty resulting from using individual AOGCMs.

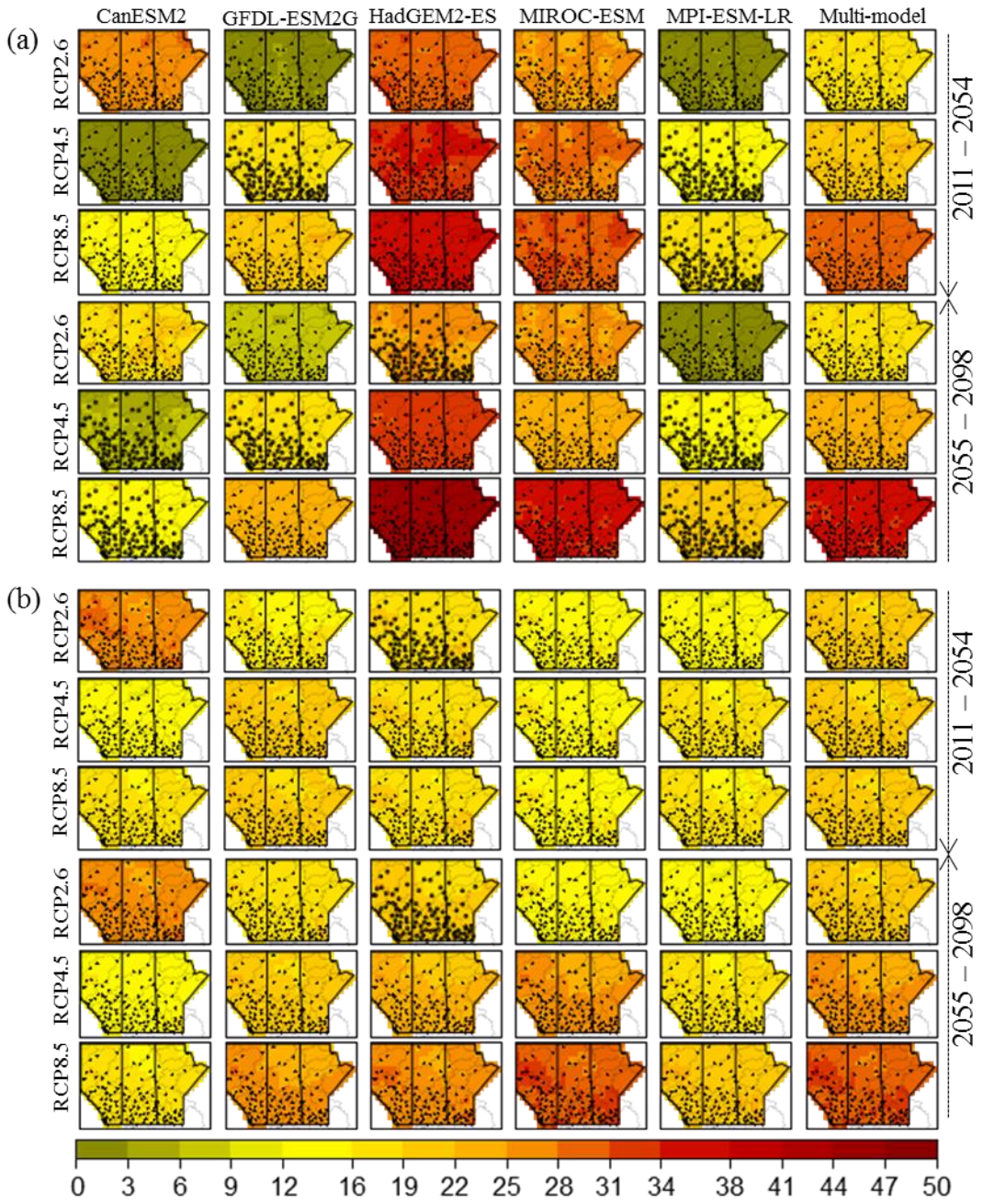

Figure 4-12: Individual and multi-model projected change (%) in 20-year return values of seasonal daily (a) minimum temperature extremes in winter and (b) maximum temperature extremes in summer. Other detail is the same as in Figure 4-9.

Most AOGCMs indicate a tendency towards largely significant changes in 20-year return levels of 3-day winter precipitation extremes across all sites in the region. In the 2030s and 2080s, the MME average projects a significant decrease in winter extremes across all sites under RCP8.5 scenario. Unlike in winter, most models indicate a tendency towards insignificant increase in summer extremes. For example, in the 2030s, the MME projects significant changes in summer extremes under RCP4.5 only while the changes are found to be insignificant in the 2080s for all scenarios. Furthermore, there exist large differences across individual AOGCMs. For example, for summer extremes, HadGEM2-ES (which projects greater changes) unlike CanESM2 projects a significant increase in summer extremes in the 2080s under all scenarios. Unlike precipitation extremes, for all scenarios, individual AOGCMs as well as the MME project largely significant changes in warm and cold temperature extremes across all sites in the region during the 2030s and 2080s.

From the findings reported so far in this study, it is found that in winter, the southern watersheds which are projected to experience the most decrease in mean precipitation also tend to experience the most declines in meteorological extremes. Contrarily, in summer, mean precipitation will probably increase more in the northern regions of the study area but precipitation extremes tend to concentrate in the southern watersheds. Also, areas with biggest extreme temperature changes are also areas with highest projected wet extremes (i.e., the southern watersheds of the study area). The projected climate changes reported here are generally consistent with those in the IPCC AR5 (IPCC, 2013), Kharin et al. (2013), and Maloney et al. (2014). For example, for projected mean climate, AR5 findings state that for the general pattern of change, "high latitudes are *very likely* to experience greater amounts of precipitation due to the increased specific humidity of the warmer troposphere as well as increased transport of water vapour from the tropics by the end of this century under the RCP8.5

scenario." Also, the largest precipitation changes over northern Eurasia and North America are projected to occur during the winter season where "it is *likely* that decreasing numbers of snowfall events are occurring where increased winter temperatures have been observed."

In terms of climate extremes, by using a multi-climate model ensemble, Kharin et al. (2013) found that precipitation and temperature extremes are projected to increase in North America. The results of their study showed cold extremes to generally warm faster than warm extremes, and more so in regions where snow and sea-ice retreat with global warming. Moreover, relative changes in the intensity of precipitation extremes will generally exceed relative changes in annual mean precipitation. It is worth noting that most of these studies examined projected climate changes on an annual scale instead of seasonal scales as performed in the current study. However, our findings agree generally with such studies which analysed climate model simulations at a coarse scale.

As mentioned previously, we downscaled only the first ensemble member from each of the six AOGCMs considered. Also, only one statistical downscaling model is applied in this study. Therefore, the conclusions based on only one downscaling model, one reanalysis product (NCEP) used to build the downscaling model, and predictors from six AOGCMs are not in any way exhaustive and may change with the inclusion of other AOGCM outputs as well as other downscaling approaches. This kind of uncertainty is a limitation of this study, but it is inherent in any statistical downscaling investigation (e.g., Katz, 2002). Certainly by addressing the limitations mentioned here, the fitted GLMs alongside other downscaling models can help improve the findings reported in this study and constitutes a potential area for future research. Finally, the validity of the assumption of stationary relationships between predictands and predictors is not verified in this study as projections of climate change refer to a future state never observed before. Many hybrid downscaling approaches (e.g., Vrac et al., 2007; Delle et al.,

2013; Sachindra et al., 2014) are currently being developed to evaluate the consistency of statistically downscaled variables.

4.5 Summary and Conclusions

This study briefly documents the performance of a multisite multivariate stochastic modelling approach based on the Generalized Linear Model (GLM) framework for simulating various characteristics of daily precipitation and minimum and maximum temperatures from 120 sites located across the Canadian Prairie Provinces of Alberta, Saskatchewan and Manitoba for the historical 1962–2005 period. The trained and tested framework is then used to downscale daily sequences of selected variables for two non-overlapping future periods (i.e., 2011–2054 and 2055–2098) using outputs from six CMIP5 AOGCMs corresponding to RCP2.6, RCP4.5, and RCP8.5 scenarios. The downscaled time series are analysed further to assess projected climate change across a rather data-sparse and hydrologically challenging region comprising 47 diverse watersheds. Based on the various analyses presented herein, the following conclusions are derived:

(1) In comparison with observations, the GLM framework simulates seasonal average values of daily precipitation and temperatures reasonably well. For precipitation, the modelling approach is able to reproduce the observed spatial patterns of daily precipitation in all seasons. For all seasons, the GLMs tend to slightly overestimate precipitation values in the middle to northern watersheds with a tendency to underestimate in the western to eastern watersheds. Overall, the simulations indicate a dry bias in winter, spring, summer, and autumn as averaged over all 120 sites. On a seasonal basis, the spatial patterns of *Tmin* are well captured in spring relative to other seasons. A slight warm bias is evident in winter, spring, and summer, while a cold bias exists in autumn. Overall, the GLMs reproduced

satisfactorily the spatial distribution of observed *Tmax*. However, the models over (under) simulated temperature values in winter and spring (summer and autumn).

(2) Concerning projected future changes in climate, relative to the historical epoch, the downscaled multi-AOGCM projections point towards an increase (decrease) in summer (winter) precipitation by the end of the 21st century. Also, in both seasons, the range of the simulated precipitation is wider for the future than the historical period meaning that future projections are more variable and uncertain, suggesting that individual AOGCM projections may be divergent. RCP8.5 scenario seems to have the larger associated variability, indicating that the sensitivity of future simulations on average is related to the magnitude of the radiative forcing. For all RCPs, *Tmin* (winter) and *Tmax* (summer) is projected to increase in the future. The increase appears to intensify with global warming and *Tmin* is expected to warm faster than *Tmax*.

(3) Winter (summer) precipitation is projected to decrease (increase) over all parts of the study area for the three RCPs by the 2030s and 2080s. Generally, the decrease (increase) is greater for the 2080s than the 2030s and will probably worsen with increased radiative forcing. For *Tmax*, in winter (summer), for the 2030s, the areal-averaged percentage change is 5 (10), 9 (12), and 14 (12.6) % while for the 2080s, this change is 15 (9), 13.8 (13), and 26 (22.5) % under the RCP2.6, RCP4.5 and RCP8.5 scenarios, respectively. The change signal is most likely to intensify in the 2080s than the 2030s and so for winter than for summer. Greatest warming is realized under the high emissions scenario (RCP8.5). Similar findings hold in the case of *Tmin*, where the greatest winter warming exceeding 40% in the 2080s under RCP8.5 in such regions as the Rocky Mountains, the northern parts of the Saskatchewan River basin, most northern watersheds and the areas to the southeast of Manitoba such as Winnipeg River watershed is likely.

(4) Projected future changes in climate extremes indices indicate that consecutive wet (dry) days will increase (decrease) with global warming. This is also true for the corresponding spell lengths. In comparison, boxplots for the 2080s are generally more dispersed than the ones for the 2030s period, meaning that the 2080s projections will probably be more variable, while RCP8.5 appears to have the largest associated variability. This variability increases with the magnitude of the GHG forcing. Similarly, hot spells are projected to increase in a warmer climate, and cold spells will decrease faster with increased warming.

(5) In terms of 20-year return levels, winter precipitation extremes are projected to decrease in future with global warming. Changes in extreme precipitation will likely exceed changes in the mean states. Conversely, precipitation extremes in summer are projected to intensify with global warming. For the 2030s (2080s), the MME projected changes in the 20-year return values of 3-day precipitation extremes 11.7 (15) % exceed those in the mean precipitation 9.4 (7) % under RCP8.5. Furthermore, warm extremes (i.e. *Tmax* maxima in summer) are projected to intensify with anthropogenic modification of the atmospheric GHG concentrations. Cold extremes (i.e. *Tmin* minima in winter) will likely warm faster than warm extremes by the end of the 21st century (2080s) under RCP8.5. Overall, the 2080s are projected to be warmer than the 2030s when compared to the historical climate.

(6) Finally, the results obtained from the multisite multivariate GLM framework for downscaling projected climate change in the Canadian Prairie Provinces are generally comparable to findings from the CMIP5 ensemble simulations. Despite some limitations of the statistical downscaling framework, it is anticipated that a trend toward wetter summers and drier winters relative to historical conditions is likely in this part of Canada. It seems that the approach adopted in this study is very flexible and could be used for other outputs and in other regions. Some limitations due to the use of a shorter ensemble can be overcome

in future studies by evaluating additional members from the CMIP5 larger ensemble. It is hoped that such endeavours will strengthen further the results of this study.

Acknowledgements

The financial support from the Global Institute for Water Security, School of Environment and Sustainability and government of Saskatchewan is gratefully acknowledged. Thanks are due to Eva Mekis from Environment Canada for providing access to adjusted precipitation and temperature data used in this study. We also thank the CMIP5 team for making available the climate model outputs. The indices of extreme events are computed using the STARDEX project FORTRAN routine.

References

Ambaum, M. H. P. (2010), Significance Tests in Climate Science, *Journal of Climate*, 23, 5927–5932. doi: http://dx.doi.org/10.1175/2010JCLI3746.1.

Arnell, N. W., B. Lloyd-Hughes (2014), The global-scale impacts of climate change on water resources and flooding under new climate and socio-economic scenarios, *Climatic Change*, 122, 1-2, 127-140, doi: 10.1007/s10584-013-0948-4.

Asong, Z.E., M.N. Khaliq, H.S. Wheater (2015), Regionalization of precipitation characteristics in the Canadian Prairie Provinces using large-scale atmospheric covariates and geophysical attributes, *Stochastic Environmental Research and Risk Assessment*, 29 (3), 875-892.

Bates, B.C., S.P. Charles, J.P. Hughes (1998), Stochastic downscaling of numerical climate model simulations, *Environmental Modelling and Software*, 13, 325–31.

Bevington, P.R., D.K. Robinson (2002), Data reduction and error analysis for the physical sciences. McGraw-Hill: New York, 320 pp.

Bonsal, B.R., R. Aider, P. Gachon, S. Lapp (2012), An assessment of Canadian prairie drought: past, present, and future, *Climate Dynamics*, 41,501–516, doi: 10.1007/s00382-012-1422-0.

Borchert, J.A. (1950), The Climate of the Central North American grassland, *Annals of the Association of American geographers*, 40, 1–39.

Chandler, R.E. (2005), On the use of generalized linear models for interpreting climate variability, *Environmetrics*, 16(7), 699–715.

Chandler, R.E. (2014), Rglimclim: A multisite, multivariate weather generator based on generalized linear models. R package version 1.3-0. http://homepages.ucl.ac.uk/~ucakarc/work/glimclim.html.

Chandler, R.E., S.M. Bate (2007), Inference for clustered data using the independence log-likelihood, *Biometrika,* 94(1), 167–183.

Chandler, R.E., H.S. Wheater (2002), Analysis of rainfall variability using generalized linear models-a case study from the West of Ireland, *Water Resources Research,* 38 (10), 1192. doi:10.1029/2001WR000906.

Charles, A., B. Timbal, E. Fernandez, H. Hendon (2013), Analog downscaling of seasonal rainfall forecasts in the Murray darling basin, *Monthly Weather Review*, 141, 1099–1117, doi: 10.1175/MWR-D-12-00098.1.

Chen, J., F. Brissette, R. Leconte, A. Caron (2012), A versatile weather generator for daily precipitation and temperature, *Transactions of the American Society of Agricultural and Biological Engineers*, 55(3), 895-906.

Chen, T.S., P.S. Yu, Y.H. Tang (2010), Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of Hydrology,* 385, 13–22.

Delle, L., F. Monache, A. Eckel, D. L. Rife, B. Nagarajan, K. Searight (2013), Probabilistic Weather Prediction with an Analog Ensemble, *Monthly Weather Review,* 141, 3498–3516.

Fischer, E. M., R. Knutti (2014), Detection of spatially aggregated changes in temperature and precipitation extremes, *Geophysical Research Letters*, 41, doi:10.1002/2013GL058499.

Fowler, H.J., S. Blenkinsopa, C. Tebaldi (2007), Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, *International Journal of Climatology*, 27, 1547–1578.

Frei, C., R. Schöll, S. Fukutome, J. Schmidli, P.L. Vidale (2006), Future change of precipitation extremes in Europe: an inter-comparison of scenarios from regional climate models, *Journal of Geophysical Research-Atmospheres*, 111, D06105, doi: 10.1029/2005JD005965.

Frost, A.J., and Coauthors (2011), A comparison of multi-site daily rainfall downscaling techniques under Australian conditions, *Journal of Hydrology,* 408,1–18.

Furrer, E. M., R.W. Katz (2007), Generalized linear modeling approach to stochastic weather generators, *Climate Research,* 34,129–144.

Goodess, C.M. (2003), Statistical and regional dynamical downscaling of extremes for European regions: STARDEX Eggs, 6, 25– 29.

Goodess, C.M., J.P. Palutikof (1998), Development of daily rainfall scenarios for southeast Spain using a circulation-type approach to downscaling, *International Journal of Climatology*, 18, 1051-1083.

Hewitson, B.C., R.G. Crane (1996), Climate downscaling: techniques and application, *Climate Research*, 7, 85–95.

Hogg, E. H., D. T. Price, T. A. Black (2000), Postulated feedbacks of deciduous forest phenology on seasonal climate patterns in the Western Canadian Interior, *Journal of Climate*, 13, 4229–4243.

IPCC (2013), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, T. F. Stocker et al. (Eds.), Cambridge University Press, Cambridge and New York, 1–1535.

Jeong, D. I., A. St-Hilaire, T.B. Ouarda, P. Gachon (20120, Multisite statistical downscaling model for daily precipitation combined by multivariate multiple linear regression and stochastic weather generator, *Climatic Change,* 114(3-4), 567- 591.

Jeong, D.I., L. Sushama, M.N. Khaliq (2014), The role of temperature in drought projections over North America, *Climatic Change,* 127, 289–303.

Jeong, D.I., L. Sushama, G.T. Diro, M.N. Khaliq, H. Beltrami, D. Caya (2015), Projected changes to high temperature events for Canada based on a regional climate model ensemble, *Climate Dynamics,* doi: 10.1007/s00382-015-2759-y.

Kalnay, E., and Coauthors (1996), The NCEP/NCAR 40-year reanalysis project, *Bulletin of American Meteorological Society*, 77(3), 437 – 471.

Katz, R.W. (1977), Precipitation as a chain-dependent process, *Journal of Applied Meteorology*, 16, 671 – 676.

Katz, R.W., M.B. Parlange, C. Tebaldi (2003), Stochastic modeling of the effects of large scale circulation on daily weather in the southeastern U.S., *Climatic Change,* 60,189–216.

Katz, R. W. (2002), Techniques for estimating uncertainty in climate change scenarios and impact studies, *Climate Research*, 20, 167–185.

Khalili, M., V.T. Nguyen, P. Gachon (2013), A statistical approach to multi-site multivariate downscaling of daily extreme temperature series, *International Journal of Climatology,* 33,15–32.

Kharin, V., F. Zwiers, X. Zhang, M. Wehner (2013), Changes in temperature and precipitation extremes in the CMIP5 ensemble, *Climatic Change*, 119(2), 345–357.

Maloney, E. D., and Coauthors (2014), North American Climate in CMIP5 Experiments: Part III: Assessment of Twenty-First-Century Projections, *Journal of Climate*, 27, 2230–2270.

Maraun, D., and Coauthors (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Reviews of Geophysics* 48, RG3003, doi:10.1029/2009RG000314.

McCullagh, P., J.A. Nelder (1989), *Generalized Linear Models*, second edition, Chapman and Hall, London, 532 pp.

Mekis, É., L.A Vincent (2011), An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada, *Atmosphere-Ocean,* 49(2), 163–177.

Moss, R.H., and Coauthors (2010), The next generation of scenarios for climate change research and assessment, *Nature*, 463(7282), 747–756.

Philips, D. (1990), Climates of Canada. Canadian Government Publishing Ottawa, Ontario, 159 pp.

Rajagopalan, B., U. Lall (1999), A k-nearest-neighbor simulator for daily precipitation and other variables, *Water Resource Research,* 35(10), 3089 – 3101.

Richardson, C.W. (1981), Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resource Research,* 17,182–190.

Richardson, C.W., D.A. Wright (1984), WGEN: A Model for Generating Daily Weather Variables. U. S. Department of Agriculture, *Agricultural Research Service,* ARS-8, 83 pp.

Sachindra, D. A., F. Huang, A. Barton, B.J.C. Perera (2012), Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment stream-flows, *International Journal of Climatology*, 33(5), 1087–1106, doi: 10.1002/joc.3493.

Sachindra, D.A., F. Huang, A.F. Barton, B.J.C. Perera (2014), Statistical downscaling of general circulation model outputs to catchment scale hydroclimatic variables, issues, challenges and possible solutions, *Journal of Water and Climate Change*, 5 (4), 496–525.

Schnur, R., D.P. Lettenmaier (1998), A case study of statistical downscaling in Australia using weather classification by recursive partitioning, *Journal of Hydrology*, 213, 362–379, doi: 10.1016/S0022-1694(98)00217-0.

Schoof, J.T., S.C. Pryor (20010, Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks, *International Journal of Climatology*, 21, 773–790.

Stern, R.D., R. Coe (1984), A model fitting analysis of daily rainfall data, *Journal of the Royal Statistical Society* Series A, 147, 1 – 34.

Taylor, K. E., R. J. Stouffer, G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, 93, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Tebaldi, C., K. Hayhoe, J.M. Arblaster, G.A. Meehl (2006), Going to the extremes; An intercomparison of model-simulated historical and future changes in extreme events, *Climatic Change*, 79, 185-211.

Todorovic, P., D.A. Woolhiser (1975), A stochastic model of n-day precipitation, *Journal of Applied Meteorology,* 14, 17 – 24.

von Storch, H., F.W. Zwiers (1999), Statistical Analysis in Climate Research, Cambridge University Press, ISBN 0521 450713, 528 pp.

Vrac, M., M. L. Stein, K. Hayhoe, X. Z. Liang (20070, A general method for validating statistical downscaling methods under future climate change, *Geophysical Research Letters*, 34, doi:10.1029/2007GL030295.

Wang, J., F. N. U. Swati, L. Michael, V. Stein, R. Kotamarthi (2015), Model performance in spatiotemporal patterns of precipitation: New methods for identifying value added by a regional climate model, *Journal of Geophysical Research: Atmospheres*, 120 (4), 1239–1259.

Wheater, H.S., P. Gober (2013), Water Security in the Canadian prairies: science and management challenges, *Philosophical Transactions of the Royal Society A*, 371, 20120409. doi: 10.1098/rsta.2012.0409.

Wheater, H.S., P. Gober (2015), Water security and the science agenda, *Water Resource Research*, Accepted Author Manuscript. doi:10.1002/2015WR016892.

Wilby, R.L., T.M.L. Wigley (1997), Downscaling general circulation model output: a review of methods and limitations, *Progress in Physical Geography*, 21, 530–548.

Wilby, R.L., S. P. Charles, E. Zorita, B. Timbal, P. Whetton, L. O. Mearns (20040, The guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change (IPCC), pre-pared on behalf of Task Group on Data and Scenario Support for Impacts and Climate Analysis (TGICA).

Wilks, D.S. (1998), Multisite generalization of a daily stochastic precipitation generation model, *Journal of Hydrology*, 210, 178–191.

Xu, C.Y. (1999), From GCMs to river flow: a review of downscaling methods and hydrologic modeling approaches, *Progress in Physical Geography*, 23 (2), 229–249.

Yang, C., R.E. Chandler, V.S. Isham, H.S. Wheater (2005), Spatial-temporal rainfall simulation using generalized linear models. *Water Resource Research*, 41, 1–13.

CHAPTER 5
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

5.1 Summary

Many research findings have shown that changes in the global climate since the 20th century were mostly due to the anthropogenic GHG emissions from human activities such as burning of fossil fuels than the natural variability of the climate. The rising GHG concentrations in the atmosphere heighten the natural greenhouse effect resulting in an imbalance in the earth's radiative energy budget leading to human induced climate change. The rising global temperatures, sea level rise due to melting glaciers and/or thermal expansion, changes in the precipitation patterns and variability which cause floods and droughts are some of the likely impacts of climate change.

Atmosphere-Ocean General Climate Models provide a reasonable basis for projecting global climate into the future. These models consider a range of scenarios ranging from GHG concentrations and other socio-economic factors for simulating global climate. Although these models are capable of providing realistic simulations of climate at global and continental scales, their coarse spatial resolution and inability to resolve sub-grid scale features such as topography, land use and cloud cover does not permit proper simulation of the climate at local, catchment and small regional scales. Consequently, direct application of the coarse resolution AOGCM outputs for catchment scale hydrological and water resources investigations is not practicable. Therefore, in order to determine local and catchment scale climate AOGCM outputs, dynamical and/or statistical downscaling techniques are commonly used.

This study was focused on improving our understanding of the historical climate and future climate change at local and catchment scales in the Canadian Prairie Provinces of Alberta, Saskatchewan and Manitoba (CPPs). Within this overall goal, specific objectives of the study

were to (1) identify large scale climate drivers of local scale climate anomalies and climate change in this region of Canada; (2) investigate the suitability of GLMs for multisite multivariate modelling of precipitation and temperature fields in the CPP region comprising 47 diverse watersheds, characterized by significant regional inhomogeneity and a paucity of ground-based observations; and (3) use the developed GLM framework to downscale Coupled Model Intercomparison Project Phase 5 (CMIP5) AOGCM outputs corresponding to three Representative Concentration Pathway (RCP) scenarios to station scale climatic variables.

The objectives of the study stated above were achieved through a case study of the CPPs. Characterized by a highly variable hydro-climate and diminishing water resources, southern parts of this region support a vibrant agro-based economy that was hard-hit by the most severe and prolonged droughts of 1988 and 1999–2005, as well as severe floods of 2011, 2013 and 2014. Hence, an analysis of the impacts of a changing climate on water resources in general and various characteristics of precipitation and temperature fields in particular is carried out in this study, which is a timely need of various socio-environmental sectors of the region.

Improved estimation of design storms by incorporating information about climate change is indispensable for use in the design, operation and maintenance of urban water infrastructure such as pipes, storm sewers, retention and detention ponds, and culverts. However, practitioners often face a common problem associated with inconsistent data sets. As a potential remedy, regional/pooled frequency analysis is utilized to improve at-site estimates or to obtain estimates at ungauged sites. The pooling process is usually based on dividing a larger region into smaller sub-regions with homogeneous characteristics of the variable of interest. However, these approaches tend to rely heavily on statistics computed from observed weather data rather than the large scale climate variables that influence the regional and local weather patterns at various

temporal and spatial scales. In this study, a new approach for identifying homogeneous regions for regionalization of precipitation characteristics was proposed for the CPPs. This approach incorporated information about large-scale atmospheric covariates, teleconnection indices and geographical site attributes in order to delineate homogeneous precipitation regions through multivariate analysis techniques – Principal Component Analysis, Canonical Correlation Analysis and Fuzzy C-means clustering. The delineated regions were validated independently for homogeneity using statistics computed from observations.

In order to drive and evaluate hydrological models for future flood and drought risk assessment at multiple points in a watershed, one often relies on observed weather data, but these data in most cases represent a major source of uncertainty in the modelling process. The most obvious case is when some observations are missing from a set of available measurements. More so, the data from different sites have different lengths and in extreme situations, inputs are required at specific ungauged locations. As an attempt to resolve these issues, missing observations are often filled in. Also, interpolation techniques have been applied to a network of weather stations to estimate data at ungauged sites. However, interpolation often leads to smoothing and reduces natural variability in observed data. As a potential solution to these issues, weather generators (WGs) have emerged. WGs are often used to generate long time-series of weather variables suitable for risk assessment. However, most of the WGs focus on individual sites and are therefore unable to represent the spatial structure of the observed weather variables. Nevertheless, for many investigations in hydrology, agriculture and environmental management, particularly in large river basins, it is important to model simultaneous sequences of multiple variables (e.g., precipitation and temperature) over large heterogeneous areas, while maintaining physically plausible spatial, temporal and inter-variable relationships. Several

approaches have been developed for simultaneous multisite multivariate generation of weather variables but they are still inadequate to model the joint distribution of, for example, precipitation and temperature simultaneously at all sites by inter-variable and inter-site dependence structures.

In this study, based on the Generalized Linear Model (GLM) framework, a multisite stochastic modelling approach was developed using daily observations of precipitation and minimum and maximum temperatures from 120 sites located across the CPPs. Temperature was modelled using a two-stage normal-heteroscedastic approach by fitting mean and variance components separately. Likewise, precipitation occurrence and conditional precipitation intensity processes were modelled separately. The relationship between precipitation and temperature was accounted for by using transformations of precipitation as covariates to predict temperature fields. NCEP Reanalysis-I large scale atmospheric covariates, teleconnection indices, geographical site attributes, and observed precipitation and temperature records were used to calibrate these models for the 1971–2000 period. Validation of the developed models was performed on both pre- and post-calibration period data.

Atmosphere Ocean General Climate Models (AOGCMs) remain the most credible tools for modelling global climate change but their very coarse spatial resolution make the outputs not readily suitable for hydrological impact studies at the local scale. Furthermore, AOGCM outputs are associated with various sources of uncertainties and also have limited or no ability to capture sub-grid scale processes which are relevant for many environmental and water resources protection studies. To bridge this gap, downscaling methods (i.e., statistical and dynamical downscaling) have often been utilized to transform AOGCM information to local and regional scale resolution. The research presented in this study explored the suitability of the GLM

framework for downscaling CMIP5 AOGCM outputs corresponding to RCP2.6, RCP4.5, and RCP8.5 scenarios to station scale weather variables over the CPPs for two non-overlapping future periods (i.e. 2011–2054 and 2055–2098).

5.2 Conclusions

The following main conclusions were drawn from this study:

(1) For determining possible large-scale forcing of the observed variability in precipitation, PCA and CCA were applied to screen NCEP-based atmospheric covariates. Apart from geographical site attributes (latitude, longitude and elevation), PCA applied to monthly precipitation totals resulted in 8 PCs that explained about two-thirds of the observed variability in precipitation. From the set of five teleconnection patterns, the indices of PDO and PNA were found to be significantly correlated with regional modes of precipitation variability over the study area. Statistically significant coherent heterogeneous canonical spatial and temporal patterns between NCEP-based atmospheric covariates and the corresponding patterns in the precipitation field were found and therefore, it can be deduced that the precipitation variability in the study area is probably linked with the variability in the large-scale atmospheric circulation at the monthly timescale. From the analysis, the significant NCEP-based atmospheric large-scale controls on local-scale precipitation in this region of Canada include 2-m air temperature, 850-hPa relative humidity, 500-hPa specific humidity, precipitable water, mean sea level pressure, horizontal wind components (850-hPa meridional and 10-m zonal wind), and vertical velocity (i.e., omega at 500-hPa).

(2) In any statistical modelling exercise, consideration of the statistical significance and consistency of the correlations between predictors and predictands of interest over time is a potential way to select a more robust set of predictors for use in a downscaling model. That is, the predictors which show correlations with fluctuations in the signs (from positive to

negative or vice versa) or largely varying magnitudes over time can be omitted, as these are indications of inconsistent relationships with the predictands. Therefore, PCA and CCA analysis in statistical modelling studies should be performed cautiously.

(3) Based on the information derived from indices of teleconnection patterns, large scale atmospheric covariates and geographic site attributes used as feature vectors in the FCM clustering algorithm, the CPPs was sub-divided into five homogeneous climatic regions: Region A–southeast of the Canadian Prairies; Region B–the northern region; Region C–the south-central region; Region D–the central west; Region E–the foothills of the Rocky Mountains to the southwest. These regions were validated for homogeneity using monthly precipitation totals as well as seasonal extremes. The identified regions were judged to be statistically as well as climatologically homogeneous. This is a significant step forward, since analysis of statistical as well as climatological homogeneity is possible now.

(4) For multisite multivariate modelling, a single model for precipitation sequences across all sites in the study area could not be realized. Therefore, separate models were developed on the basis of five pre-defined homogeneous regions covering the study area. Consequently, a joint model for both precipitation and temperature variables was not feasible. As an alternative, precipitation from different homogeneous regions was used as additional covariates to predict temperature. The results from both calibration and validation periods showed very good agreement between simulated and observed values of various precipitation and temperature characteristics (such as mean, autocorrelation, wet-day proportions as well as inter-annual variability and spatial dependence structure) for each month of the year although better accuracy was realized in the case of temperature relative to precipitation. So far, it was difficult to find any plausible WG which has been applied at

such a spatial scale with that skill. Therefore, one can conclude that the GLM framework can reliably be used for downscaling weather variables in this region of Canada.

(5) A robust framework to downscale local daily weather series from coarse AOGCM outputs was developed and tested over several sites located across the study area. The GLM approach employed here is probabilistic in nature and is simple to apply for simulating multiple weather sequences which is a clear advantage over RCMs because it produces realistic and unbiased point-scale weather data. Also, given that a multi-GHG scenario multi-AOGCM approach was used in this study, the GLM probabilistic framework allows for generating a large number of simulations to handle uncertainties. For example, it was found here that the sensitivity of future simulations is probably a function of the magnitude of the GHG forcing scenario and there is substantial variation from one AOGCM to the other. The GLM framework is therefore useful for driving impact studies that focus on daily to seasonal scales in this region of Canada.

(6) The downscaled results indicated that winter (summer) mean precipitation is projected to decrease (increase) over all parts of the study area for the three RCPs by the 2030s and 2080s. Generally, the decrease (increase) is greater for the 2080s than the 2030s and will probably worsen with increased radiative forcing. For instance, in winter, the MME areal-averaged (over 120 sites) change in precipitation in the 2030s (2080s) is -4 (-4.8), -3 (-12), and -10 (-23) % under RCP2.6, RCP4.5 and RCP8.5 scenarios, respectively. While in summer, the MME projected increase in mean precipitation over all sites in the 2030s (2080s) is 10.3 (9.8), 9 (8.2), and 9.4 (7) %, respectively for RCP2.6, RCP4.5 and RCP8.5 scenarios. Concerning regional patterns of change, the least increases are expected to occur over southern regions pertaining to the Lake Winnipeg, Lake Manitoba, Athabasca River, and North Saskatchewan River watersheds. This signal is consistent over all scenarios and

temporal periods. For *Tmax*, in winter (summer), for the 2030s, the areal-averaged percentage change is 5 (10), 9 (12), and 14 (12.6) % while for the 2080s, this change is 15 (9), 13.8 (13), and 26 (22.5) % under the RCP2.6, RCP4.5 and RCP8.5 scenarios, respectively. The change signal is most likely to intensify in the 2080s than the 2030s and so for winter than for summer. Greatest warming is realized under the high emissions scenario (RCP8.5). Similar findings hold in the case of *Tmin*, where the greatest winter warming exceeding 40% in the 2080s under RCP8.5 in such regions as the Rocky Mountains, the northern parts of the Saskatchewan River basin, most northern watersheds and the areas to the southeast of Manitoba such as Winnipeg River watershed is likely.

(7) Changes in seasonal precipitation and temperature extremes were also evaluated in this study using extreme value theory. In terms of 20-year return levels, winter precipitation extremes are projected to decrease in future with global warming. For example, the areal-averaged MME projected decrease in 20-year return levels over all sites in the 2030s (2080s) is 4.3 (4.8), 5.4 (6.2), and 7.4 (-28) %, respectively for RCP2.6, RCP4.5 and RCP8.5 scenarios. The response is more pronounced for the 2080s compared to the 2030s.Changes in extreme precipitation will likely exceed changes in the mean states. Conversely, precipitation extremes in summer are projected to intensify with global warming. For the 2030s (2080s), the MME projected changes in the 20-year return values of 3-day precipitation extremes 11.7 (15) % exceed those in the mean precipitation 9.4 (7) % under RCP8.5. Furthermore, warm extremes (i.e. *Tmax* maxima in summer) are projected to intensify with anthropogenic modification of the atmospheric GHG concentrations. Cold extremes (i.e. *Tmin* minima in winter) will likely warm faster than warm extremes by the end of the 21st century (2080s) under RCP8.5. The areal-averaged MME projected increase in 20-year *Tmin* return levels in the 2030s (2080s) is 18 (18.8), 21 (23), and 29 (34) %, respectively for RCP2.6, RCP4.5 and

RCP8.5 scenarios. While the areal-averaged MME projected increase in 20-year *Tmax* quantiles in the 2030s (2080s) is 15 (17), 17 (21), and 24 (30) %, respectively for RCP2.6, RCP4.5 and RCP8.5 scenarios. Overall, the 2080s are projected to be warmer than the 2030s when compared to the historical climate (1962 – 2005).

5.3 Recommendations for Future Work

The analyses presented in this study are in no way exhaustive and can be improved and expanded in various ways. The following recommendations can be exploited in the future to improve on or complement these analyses:

(1) The regionalization approach proposed in this study for delineating homogeneous precipitation regions can potentially be improved in future studies. The analyses reported here have shown that a target region of interest can be subdivided into smaller homogeneous regions based on large-scale atmospheric covariates, indices of teleconnection patterns and geographic site attributes that control weather processes at local and regional scales. It is recommended that future studies should also investigate other attributes such as the physiographic characteristics of watersheds and indices of temperature.

(2) In this study, the RFA algorithm was applied to defuzzified homogeneous soft regions to avoid issues related to the presence of sites with fractional memberships in various regions due to lack of a theoretical framework. Defuzzification can lead to a loss of significant information such as inter-regional correlations within a multisite multivariate downscaling framework which can make it difficult to capture aspects of weather extremes that traverse regions. Further work should aim at nesting homogeneous soft regions and developing a sound theoretical framework.

(3) Here, only the first run from each of the six AOGCMs considered was downscaled. Also, only one statistical downscaling model was applied. Thus, conclusions based only on one

downscaling model, one reanalysis product (NCEP) used to build the downscaling model, and predictors from six AOGCMs are not in any way optimal and may change with inclusion of other AOGCM outputs as well as downscaling models.

(4) Moreover, as was the case in this study, it is a common practice to train and test statistical downscaling models with reanalysis outputs and then use AOGCM outputs for developing projections at the station scale. Given that the inputs used in the development and future projection phases of the downscaling models originate from two different sources, inhomogeneous behavior of the resulting sequences is not surprising. This procedure in most cases does not provide a smooth transition from the downscaling model development phase to the future projection phase. Further work is needed in order to address this issue adequately.

(5) Finally, the validity of the statistical downscaling assumption of stationary relationships between predictands and predictors was not verified in this study as projections of climate change refer to a future state never observed before. Many hybrid downscaling approaches are currently being developed to evaluate the consistency of statistically downscaled variables. This also constitutes another area for future research.

APPENDIX A: ATTRIBUTES OF OBSERVATION STATIONS

Table A-1: Attributes of observation stations used in the study. The homogeneous region associated with each station is also indicated. The sites where temperature records are not available are shown in bold.

ID	Site	Station name	Region	Eastings	Northings	Elevation (m)
1	GG89	HUDSON BAY	A	-102.58	52.88	422
2	GG90	PELLY	A	-101.87	52.08	509
3	GG93	CYPRESS RIVER	A	-99.08	49.55	374
4	**GG95**	PORTAGE PRAIRIE	A	-98.27	49.95	259
5	**GG96**	EMERSON AUT	A	-97.23	49.00	242
6	GG98	MORDEN	A	-98.08	49.18	298
7	GG99	SPRAGUE	A	-95.60	49.02	329
8	**G100**	STEINBACH	A	-96.77	49.53	254
9	G101	WINNIPEG	A	-97.23	49.92	239
10	G102	ARBORG	A	-97.08	50.93	224
11	G103	BERENS RIVER	A	-97.03	52.35	222
12	**G104**	BISSETT	A	-95.70	51.03	259
13	G105	GIMLI	A	-97.02	50.63	223
14	G106	GRAND RAPIDS	A	-99.28	53.15	223
15	G107	GREAT FALLS	A	-96.00	50.47	249
16	G108	INDIAN BAY	A	-95.20	49.62	327
17	G109	PINAWA WNRE	A	-96.07	50.18	267
18	G110	DAUPHIN	A	-100.05	51.10	305
19	**G111**	SWAN RIVER	A	-101.23	52.12	335
20	**G112**	LANGRUTH WEST	A	-98.80	50.42	264
21	G113	NEEPAWA MURRAY	A	-99.57	50.15	412
22	G115	THE PAS	A	-101.10	53.97	270
23	G119	NORWAY HOUSE	A	-97.85	53.97	224
24	GG35	FORT MCMURRAY	B	-111.22	56.65	369
25	GG40	FORT CHIPEWYAN	B	-111.12	58.77	232
26	GG41	FORT VERMILION	B	-116.03	58.38	289
27	GG43	KEG RIVER RS	B	-117.62	57.75	405
28	GG79	URANIUM CITY	B	-108.48	59.57	318
29	GG80	COLLINS BAY	B	-103.70	58.18	490
30	**GG81**	CREE LAKE	B	-107.13	57.35	495
31	GG82	ISLAND FALLS	B	-102.35	55.53	299
32	GG83	KEY LAKE	B	-105.62	57.25	509
33	GG84	LA RONGE	B	-105.27	55.15	379
34	GG87	WHITESAND DAM	B	-103.15	56.23	344
35	G114	FLIN FLON	B	-101.88	54.77	320
36	G116	CHURCHILL	B	-94.07	58.73	29
37	G117	GILLAM	B	-94.72	56.35	145
38	G118	LYNN LAKE	B	-101.08	56.87	357
39	G120	THOMPSON	B	-97.87	55.80	222
40	GGG3	CORONATION	C	-111.45	52.07	791
41	GG11	JENNER	C	-111.20	50.72	755
42	GG14	SCOTFIELD	C	-111.35	51.58	762
43	GG24	ONEFOUR	C	-110.47	49.12	935
44	**GG47**	BANGOR	C	-102.28	50.90	526
45	**GG48**	CEYLON	C	-104.65	49.38	753
46	GG49	COTE	C	-101.78	51.52	450
47	GG50	DAVIDSON	C	-105.98	51.27	619
48	GG51	ESTEVAN	C	-102.97	49.22	581
49	GG52	INDIAN HEAD	C	-103.65	50.55	579
50	GG53	KELLIHER	C	-103.75	51.25	676
51	**GG54**	MANOR	C	-102.10	49.62	633
52	GG55	MOOSE JAW	C	-105.55	50.33	577
53	GG56	MOOSOMIN	C	-101.67	50.13	576
54	GG57	PASWEGIN	C	-103.92	51.98	533
55	GG58	REGINA	C	-104.67	50.43	577

56	GG59	YELLOW GRASS	C	-104.18	49.82	580
57	GG60	TONKIN	C	-102.23	51.20	527
58	GG61	ANEROID	C	-107.30	49.72	754
59	GG62	LEADER	C	-109.50	50.90	676
60	**GG63**	CHAPLIN	C	-106.65	50.47	672
61	**GG64**	HIGH POINT	C	-107.93	50.98	645
62	**GG65**	KLINTONEL	C	-108.92	49.68	1074
63	GG66	SWIFT CURRENT	C	-107.73	50.27	825
64	GG67	VAL-MARIE	C	-107.85	49.37	808
65	GG68	WEST POPLAR	C	-106.38	49.00	876
66	GG69	KINDERSLEY	C	-109.18	51.52	694
67	GG70	BATTLEFORD	C	-108.25	52.77	548
68	GG71	SCOTT	C	-108.83	52.37	660
69	GG72	WASECA	C	-109.40	53.13	638
70	**GG73**	MELFORT	C	-104.60	52.82	490
71	GG74	OUTLOOK	C	-107.05	51.48	541
72	GG75	PILGER	C	-105.15	52.42	552
73	GG76	PRINCE ALBERT	C	-105.67	53.22	428
74	GG77	SASKATOON	C	-106.72	52.17	504
75	GG86	WASKESIU LAKE	C	-106.07	53.92	569
76	GG88	NIPAWIN	C	-104.00	53.33	372
77	GG91	BIRTLE	C	-101.05	50.43	522
78	GG92	BRANDON	C	-99.95	49.92	409
79	GG94	PIERSON	C	-101.27	49.18	469
80	**GG97**	NINETTE	C	-99.65	49.42	419
81	GGG1	CALMAR	D	-113.85	53.28	720
82	GGG2	CAMROSE	D	-112.82	53.03	739
83	GGG4	EDMONTON	D	-113.58	53.32	723
84	**GGG5**	ELK POINT	D	-111.07	53.88	605
85	GGG6	RANFURLY	D	-111.73	53.42	673
86	GGG8	SION	D	-114.12	53.88	701
87	GG30	ATHABASCA	D	-113.28	54.72	515
88	GG32	CAMPSIE	D	-114.68	54.13	671
89	**GG34**	ENILDA-BERG	D	-116.30	55.42	591
90	GG36	SLAVE LAKE	D	-114.78	55.28	583
91	GG37	WHITECOURT	D	-115.78	54.15	782
92	GG38	BEAVERLODGE	D	-119.40	55.20	745
93	GG39	FAIRVIEW	D	-118.53	56.08	604
94	GG42	GRANDE PRAIRIE	D	-118.88	55.18	669
95	GG44	PEACE RIVER	D	-117.45	56.23	571
96	**GG45**	WABASCA RS	D	-113.83	55.97	545
97	GG46	COLD LAKE	D	-110.28	54.42	541
98	GG78	BUFFALO NARROWS	D	-108.43	55.83	440
99	GG85	LOON LAKE	D	-109.10	54.05	543
100	GGG7	ROCKY MT HOUSE	E	-114.92	52.42	988
101	GGG9	STETTLER NORTH	E	-112.72	52.33	821
102	**GG10**	DRUMHELLER	E	-112.87	51.47	719
103	GG12	LACOMBE 2	E	-113.75	52.45	860
104	**GG13**	OLDS	E	-114.10	51.78	1040
105	GG15	CALGARY	E	-114.02	51.12	1084
106	GG16	CLARESHOLM	E	-113.73	49.93	1035
107	GG17	CARWAY	E	-113.38	49.00	1354
108	GG18	GLEICHEN	E	-113.05	50.88	905
109	GG19	LETHBRIDGE	E	-112.80	49.63	929
110	GG20	MEDICINE HAT	E	-110.72	50.02	717
111	**GG21**	MOUNTAIN VIEW	E	-113.63	49.13	1339
112	GG22	PINCHER CREEK	E	-113.98	49.52	1190
113	GG23	VAUXHALL	E	-112.13	50.05	779
114	GG25	BANFF	E	-115.55	51.20	1397
115	**GG26**	BEAVER MINES	E	-114.18	49.47	1257
116	**GG27**	CROWSNEST	E	-114.48	49.63	1303
117	**GG28**	HIGHWOOD	E	-114.37	50.55	1580
118	GG29	JASPER WARDEN	E	-118.03	52.93	1020
119	GG31	HINTON VALLEY	E	-117.53	53.40	1011
120	GG33	EDSON	E	-116.45	53.58	927

APPENDIX B: FITTED GLM SPECIFICATIONS

In the case of precipitation, model specifications as well as residual plots are presented for homogeneous Region C to give a general picture of the fitted models using data for the period 1971–2000. For temperature, all model specifications and residual plots are provided.

B1 Summary of occurrence and amounts models for daily precipitation

B1.1 Occurrence model

Main effects	Coefficient	Std Error
Constant	-2.0006	0.0469
Legendre polynomial 1 for Eastings (Longitude)	0.5309	0.0425
Legendre polynomial 1 for Northings (Latitude)	0.4255	0.0380
Legendre polynomial 2 for Elevation(m)	-0.1875	0.0474
Pacific North America Oscillation	-0.4253	0.0533
Daily seasonal effect, cosine component	-1.2928	0.0812
Daily seasonal effect, sine component	-0.4129	0.0326
Mean of Ln(1+Precipitation[t-1])	1.0889	0.0641
Mean of Ln(1+Precipitation[t-2])	-0.4337	0.0336
Mean of Ln(1+Precipitation[t-3])	0.2027	0.0299
Mean of I(Precipitation[t-k]>0: k=1 to 2)	0.2718	0.0184
2-m air temperature	0.7255	0.0684
850-hPa relative humidity	0.2768	0.0172
Mean sea level pressure	-0.3247	0.0197
850-hPa Uwind	-0.3055	0.0159
10-m Vwind	0.2040	0.0157

Two-way interactions:	Coefficient	Std Error
Daily seasonal effect, cosine component with 2-m air temperature	1.0975	0.0655
Daily seasonal effect, sine component with 2-m air temperature	0.5476	0.0527
2-m air temperature with Mean of Ln(1+Precipitation[t-1])	-0.3994	0.0647

B1.2 Amounts model

Main effects	Coefficient	Std Error
Constant	2.6590	0.0538
Legendre polynomial 1 for Eastings (Longitude)	0.1605	0.0445
Legendre polynomial 1 for Northings (Latitude)	0.0231	0.0073
Arctan-transformed Elevation(m)	-0.7448	0.0636
Daily seasonal effect, cosine component	0.0845	0.0214
Daily seasonal effect, sine component	0.342	0.0189
Mean of Ln(1+Precip[t-1])	0.2231	0.0698
Mean of Ln(1+Precip[t-2])	0.0975	0.0278
Mean of I(Precip[t-k]>0: k=1 to 2)	-0.3548	0.0223
2-m air temperature	0.1172	0.0260
850-hPa relative humidity	0.0938	0.0200
850-hPa Uwind	0.5496	0.0588
10-m Vwind	0.1362	0.0173

Two-way interactions:	Coefficient	Std Error
2-m air temperature with Daily seasonal effect, cosine component	-0.1931	0.0271
2-m air temperature with Daily seasonal effect, sine component	-0.1022	0.0272
Daily seasonal effect, sine component With Mean of Ln(1+Precip[t-1])	-1.3280	0.0288
Daily seasonal effect, cosine component with Mean of Ln(1+Precip[t-1])	0.2569	0.0862

B1.3 Pearson residual plots for fitted GLMs

B1.3.1 Occurrence model — monthly and annual residuals

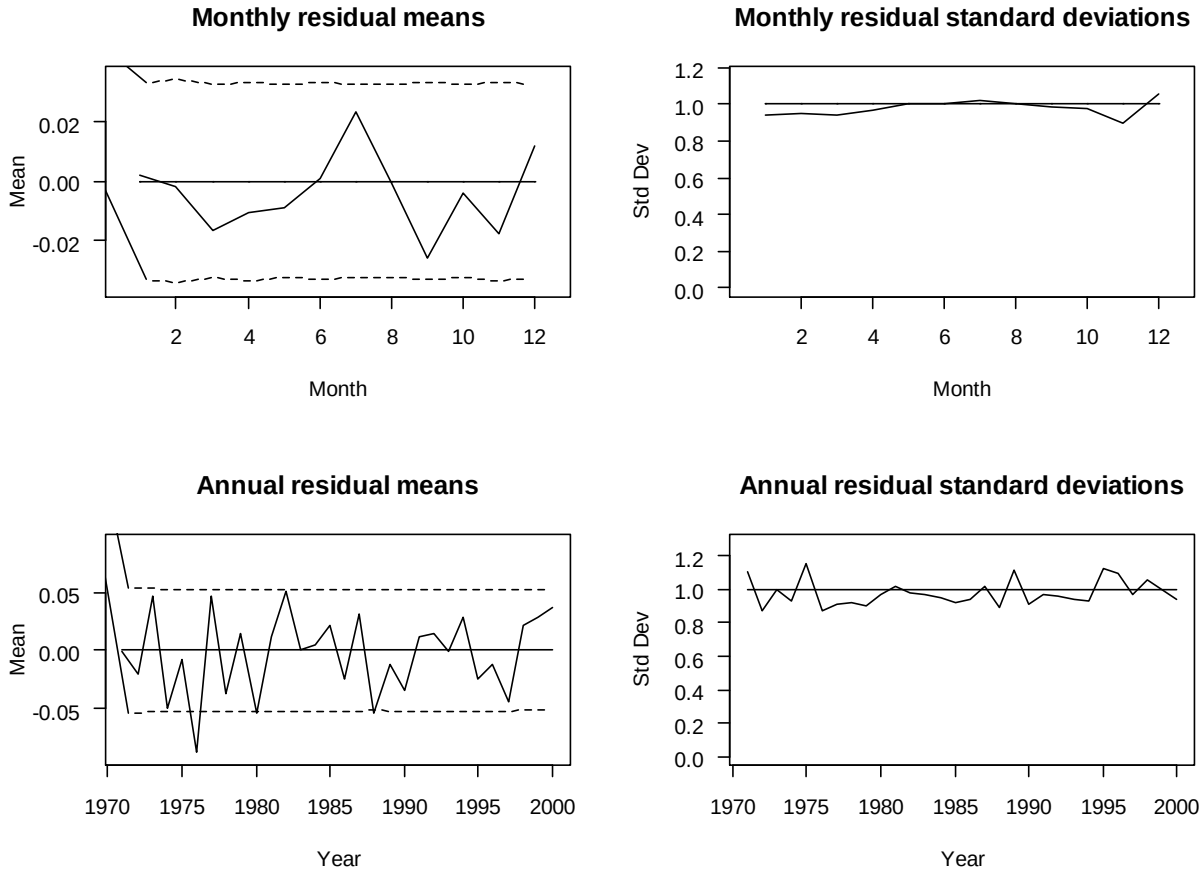

Figure B1: Pearson residuals for the occurrence model. Means and standard deviations are shown by month and year. Dashed lines in the plots on the left correspond to 95% confidence bands under the fitted model.

B1.3.2 Amounts model — residuals by month and year

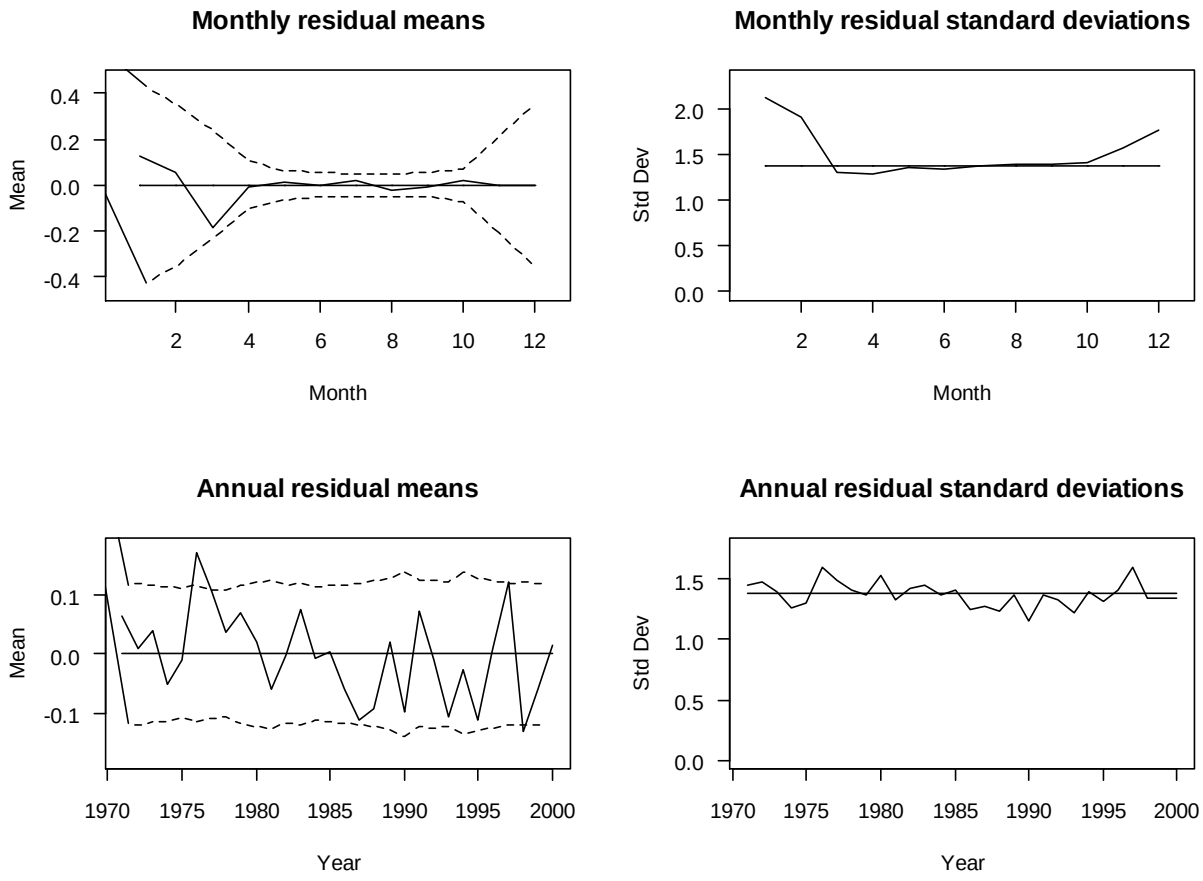

Figure B2: Pearson residuals for the amounts model. Means and standard deviations are shown by month and year. Other information is the same as in Figure B1.

B1.3.3 Occurrence model — residuals by site

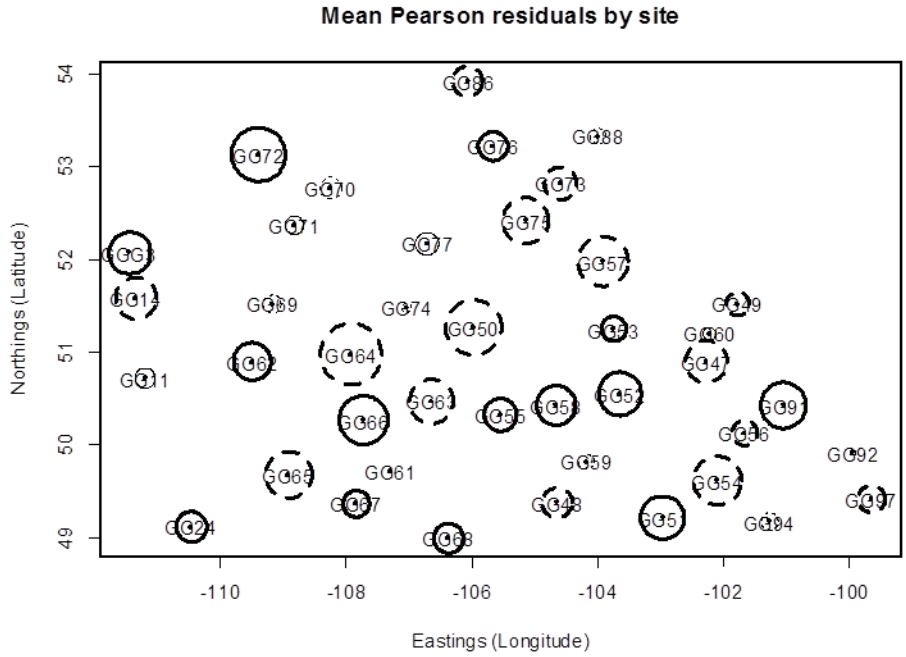

Figure B3: Bubble map showing mean residuals from the fitted occurrence model at each site in Region C. Circle areas are proportional to standardized mean residuals. There is no discernible spatial organization of negative (dashed circles) and positive (solid circles) residuals. Site codes are enclosed within circles and are elaborated in Appendix A.

B1.3.4 Amounts model — residuals by site

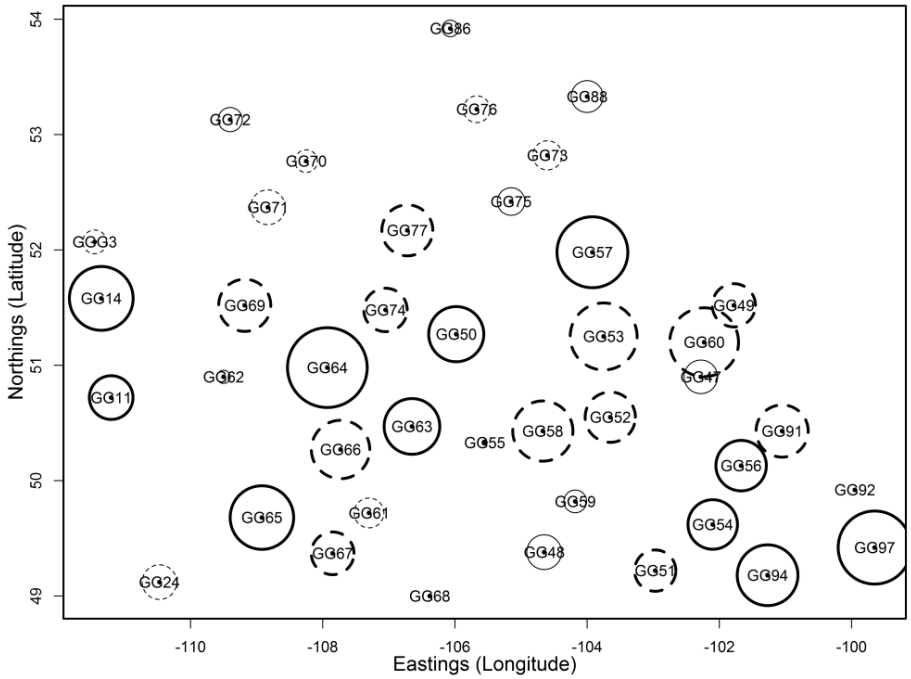

Figure B4: Bubble map showing mean residuals from the fitted amounts model at each site in Region C. Circle areas are proportional to standardized mean residuals. There is no discernible spatial organization of negative (dashed circles) and positive (solid circles) residuals.

B2 Summary of jointly fitted mean and variance model for daily minimum temperature

Response variable: Minimum temperature

Mean structure

Main effects:	Coefficient	Std Err
Constant	-1.0941	0.0397
Legendre polynomial 1 for Eastings (Longitude)	-0.0242	0.0279
Legendre polynomial 1 for Northings (Latitude)	-0.1874	0.0320
Legendre polynomial 1 for Elevation(m)	-0.2269	0.0306
Daily seasonal effect, cosine component	0.6955	0.0654
Daily seasonal effect, sine component	0.5382	0.0316
Tmin[t-1]	0.3175	0.0025
Tmin[t-2]	0.0857	0.0024
RA Mean of Ln(1+Precip[t])	0.4603	0.0395
RB Mean of Ln(1+Precip[t])	0.2975	0.0289
RC Mean of Ln(1+Precip[t])	0.0450	0.0280
RD Mean of Ln(1+Precip[t])	0.2643	0.0202
RE Mean of Ln(1+Precip[t])	0.3513	0.0185
RE Mean of Ln(1+Precip[t-1])	-0.2418	0.0388
2-m air temperature	0.2137	0.0658
850-hPa relative humidity	0.3389	0.0138
500-hPa specific humidity	0.4685	0.0179
Mean sea level pressure	0.2862	0.0170
850-hPa Uwind	-0.0464	0.0131
10-m Vwind	0.5891	0.0132

2-way interactions:	Coefficient	Std Err
Daily seasonal effect, cosine component with Tmin[t-1]	0.0776	0.0024
Daily seasonal effect, sine component with Tmin[t-1]	0.0376	0.0024
Daily seasonal effect, cosine component with Legendre polynomial 1 for Northings (Latitude)	-0.1119	0.0350
Daily seasonal effect, sine component with Legendre polynomial 1 for Northings (Latitude)	-0.3463	0.0327
Legendre polynomial 1 for Elevation(m) with Daily seasonal effect, cosine component	0.2337	0.0451
Legendre polynomial 1 for Elevation(m) with Daily seasonal effect, sine component	0.2948	0.0386

Dispersion structure:

Main effects:	Coefficient	Std Err
Constant	2.7585	0.0135
Fourier cosine component 1 for Eastings (Longitude)	0.0909	0.0040
Legendre polynomial 2 for Northings (Latitude)	0.1539	0.0081
Arctan-transformed Elevation(m)	-0.4313	0.0188
Pacific Decadal Oscillation	-0.2314	0.0039
Daily seasonal effect, cosine component	0.1260	0.0216
Daily seasonal effect, sine component	0.2102	0.0117
Tmin[t-1]	-0.0104	0.0009
Tmin[t-2]	-0.0045	0.0008
RE Mean of Ln(1+Precip[t])	-0.9521	0.0168
RE Mean of Ln(1+Precip[t-2])	-0.0343	0.0168
2-m air temperature	-0.2322	0.0212
850-hPa relative humidity	0.0379	0.0047
500-hPa specific humidity	0.1642	0.0069
850-hPa Uwind	0.0562	0.0045

2-way interactions:	Coefficient	Std Err
Legendre polynomial 2 for Northings (Latitude) with Daily seasonal effect, cosine component	-0.1300	0.0115
Legendre polynomial 2 for Northings (Latitude) with Daily seasonal effect, sine component	-0.1034	0.0114
Arctan-transformed Elevation(m) with Daily seasonal effect, cosine component	0.1900	0.0275
Arctan-transformed Elevation(m) with Daily seasonal effect, sine component	-0.4523	0.0258
Daily seasonal effect, cosine component with Tmin[t-1]	0.0110	0.0009
Daily seasonal effect, sine component with Tmin[t-1]	-0.2843	0.0009

B2.1 Jointly fitted mean and variance model — residuals by month and year

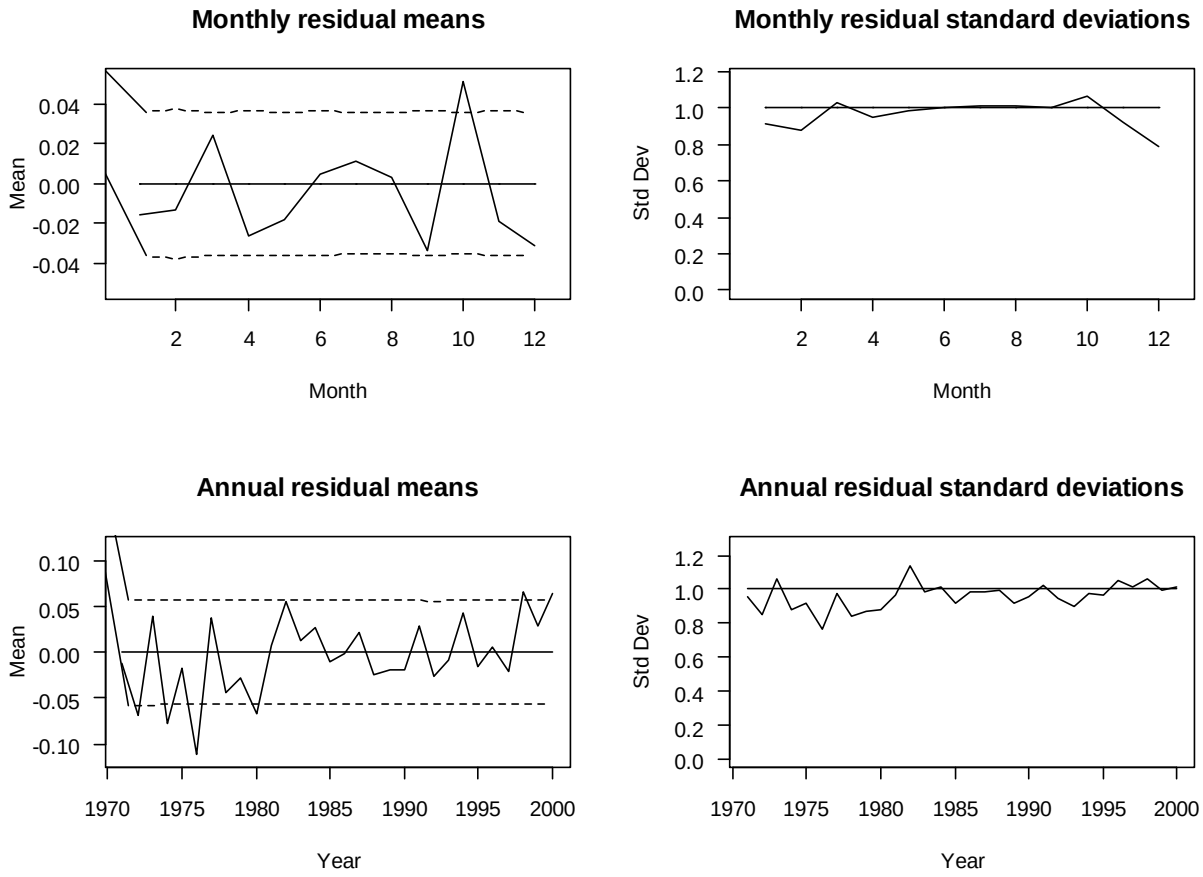

Figure B5: Pearson residuals for the jointly fitted mean and variance model for minimum temperature. Means and standard deviations are shown by month and year. Dashed lines in the plots on the left correspond to 95% confidence bands under the fitted model.

B2.2 Jointly fitted mean and variance model — residuals by site

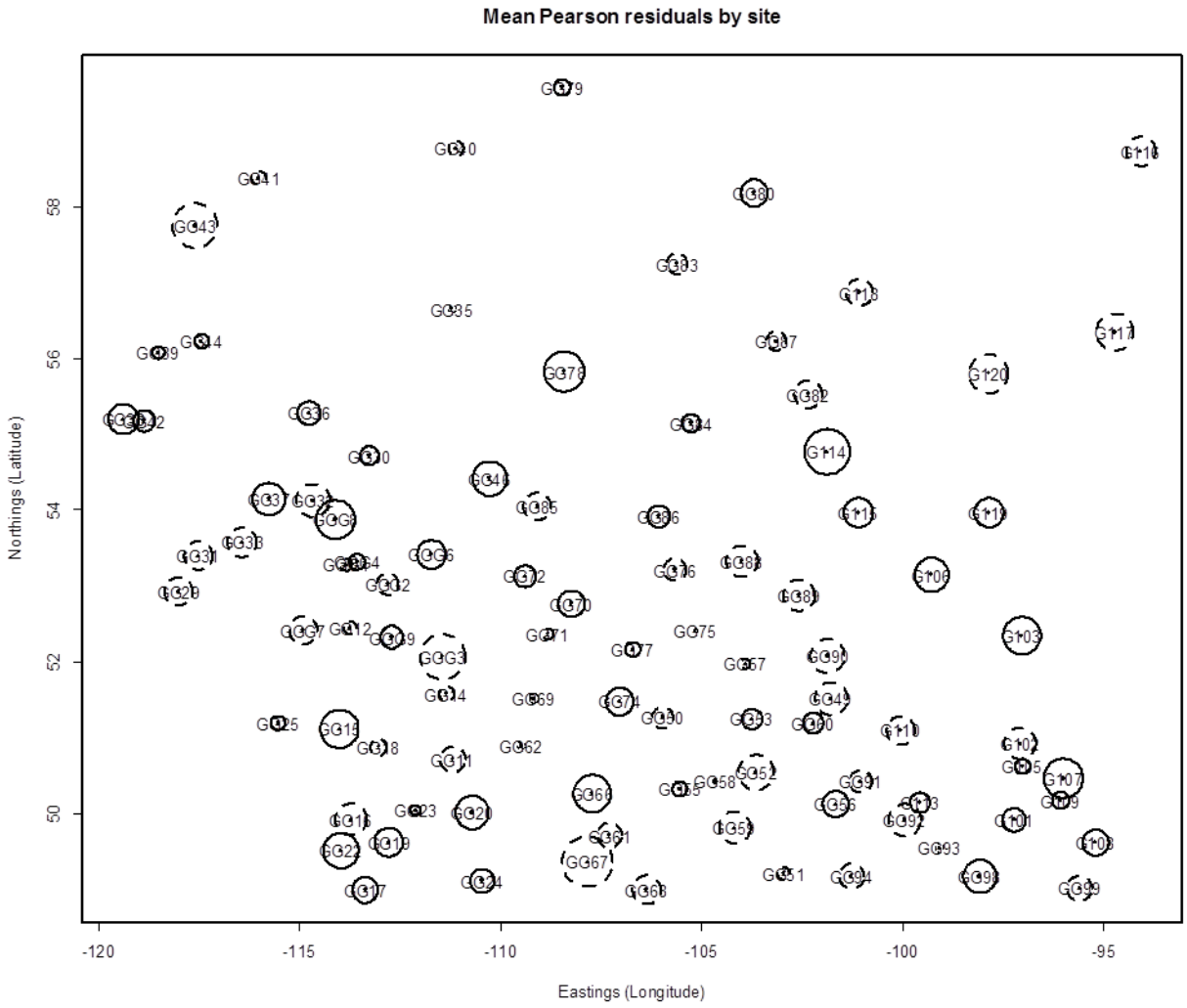

Figure B6: Bubble map showing mean residuals from the jointly fitted mean and variance model at each site in the entire study region. Circle areas are proportional to standardized mean residuals. There is no discernible spatial organization of negative (dashed circles) and positive (solid circles) residuals.

B3 Summary of jointly fitted mean and variance model for daily maximum temperature

Response variable: Maximum temperature

Mean structure

Main effects:	Coefficient	Std Err
Constant	3.1309	0.0445
Legendre polynomial 1 for Eastings (Longitude)	-0.2792	0.0368
Legendre polynomial 1 for Northings (Latitude)	-0.3210	0.0383
Legendre polynomial 1 for Elevation(m)	0.6816	0.0341
Monthly half-year cycle, cosine component	0.0512	0.0311
Monthly half-year cycle, sine component	-0.2223	0.0330
Daily seasonal effect, cosine component	0.2156	0.0904
Daily seasonal effect, sine component	0.1601	0.0418
Distance-weighted mean of Tmax[t-1]	0.5911	0.0043
Tmin[t]	0.9093	0.0042
RE Mean of Ln(1+Precip[t])	0.6771	0.0664
2-m air temperature	0.9158	0.0823
850-hPa relative humidity	-0.4278	0.0189
Mean sea level pressure	0.4751	0.0232
850-hPa Uwind	0.7932	0.0195
10-m Vwind	0.8020	0.0187

2-way interactions:	Coefficient	Std Err
Daily seasonal effect, cosine component with Distance-weighted mean of Tmax[t-1]	0.1050	0.0045
Daily seasonal effect, sine component with Distance-weighted mean of Tmax[t-1]	0.2501	0.0044
Daily seasonal effect, cosine component with 2-m air temperature	-0.4351	0.0046
Daily seasonal effect, sine component with 2-m air temperature	-0.1104	0.0045

Dispersion structure:

Main effects:	Coefficient	Std Err
Constant	1.9423	0.0277
Legendre polynomial 1 for Eastings (Longitude)	0.1637	0.0052
Legendre polynomial 2 for Northings (Latitude)	0.2326	0.0101
Arctan-transformed Elevation(m)	0.4515	0.0256
PDO	-0.2742	0.0047
Daily seasonal effect, cosine component	0.2723	0.0253
Daily seasonal effect, sine component	0.0704	0.0120
Distance-weighted mean of Tmax[t-1]	0.1750	0.0028
RA Mean of Ln(1+Precip[t])	0.3764	0.0954
RB Mean of Ln(1+Precip[t])	0.2475	0.0221
RC Mean of Ln(1+Precip[t])	0.1450	0.0243
RD Mean of Ln(1+Precip[t])	0.2665	0.0114
RE Mean of Ln(1+Precip[t])	0.4232	0.0034
2-m air temperature	0.3077	0.0105
850-hPa relative humidity	-1.2619	0.0342
850-hPa Uwind	0.0441	0.0066

2-way interactions:	Coefficient	Std Err
Legendre polynomial 2 for Northings (Latitude) with Daily seasonal effect, cosine component	-0.1321	0.0512
Legendre polynomial 2 for Northings (Latitude) with Daily seasonal effect, sine component	-0.1354	0.0273
Arctan-transformed Elevation(m) with Daily seasonal effect, cosine component	0.1823	0.0275
Arctan-transformed Elevation(m) with Daily seasonal effect, sine component	-0.4523	0.0258
Daily seasonal effect, cosine component with Distance-weighted mean of Tmax[t-1]	-0.0122	0.0008
Daily seasonal effect, sine component with Distance-weighted mean of Tmax[t-1]	0.3427	0.0007

B3.1 Jointly fitted mean and variance model — residuals by month and year

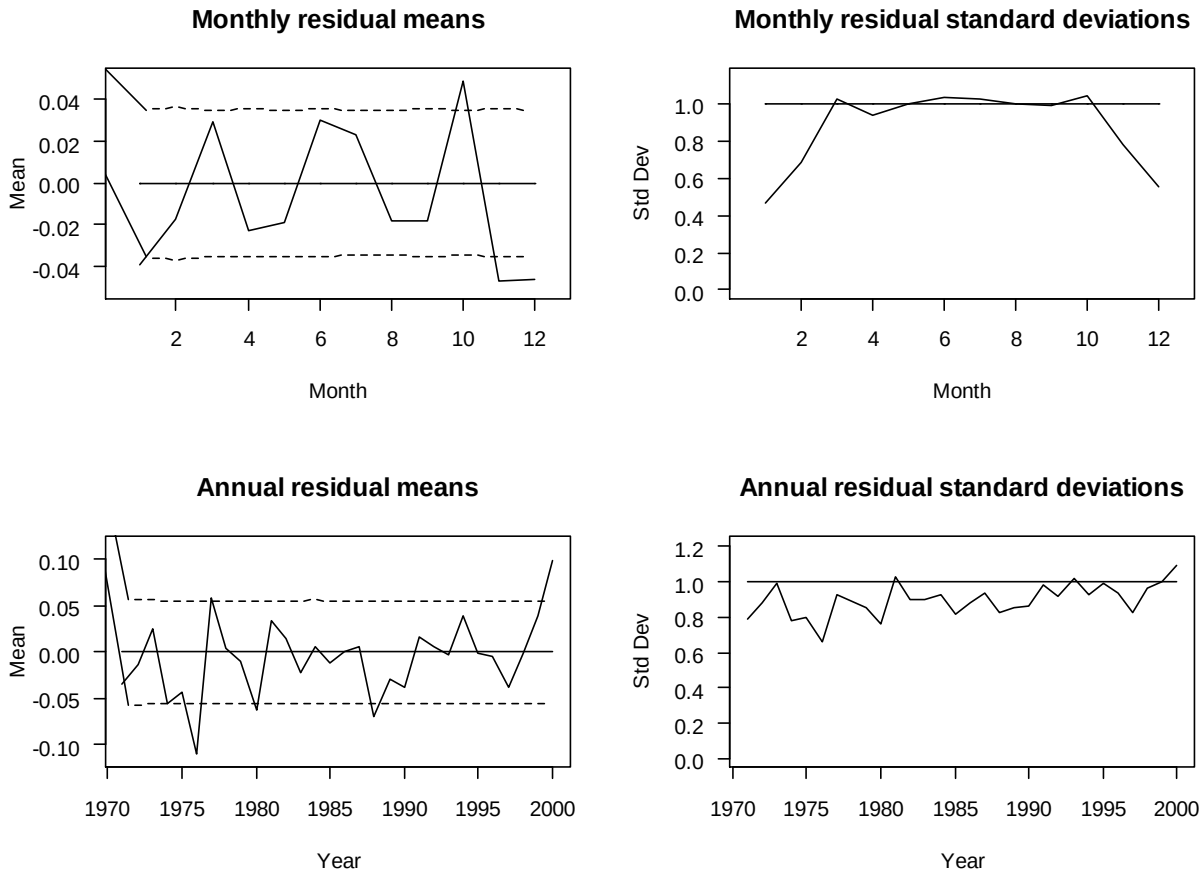

Figure B7: Pearson residuals for the jointly fitted mean and variance model for maximum temperature. Means and standard deviations are shown by month and year. Dashed lines in the plots on the left correspond to 95% confidence bands under the fitted model.

B3.2 Jointly fitted mean and variance model – residuals by site

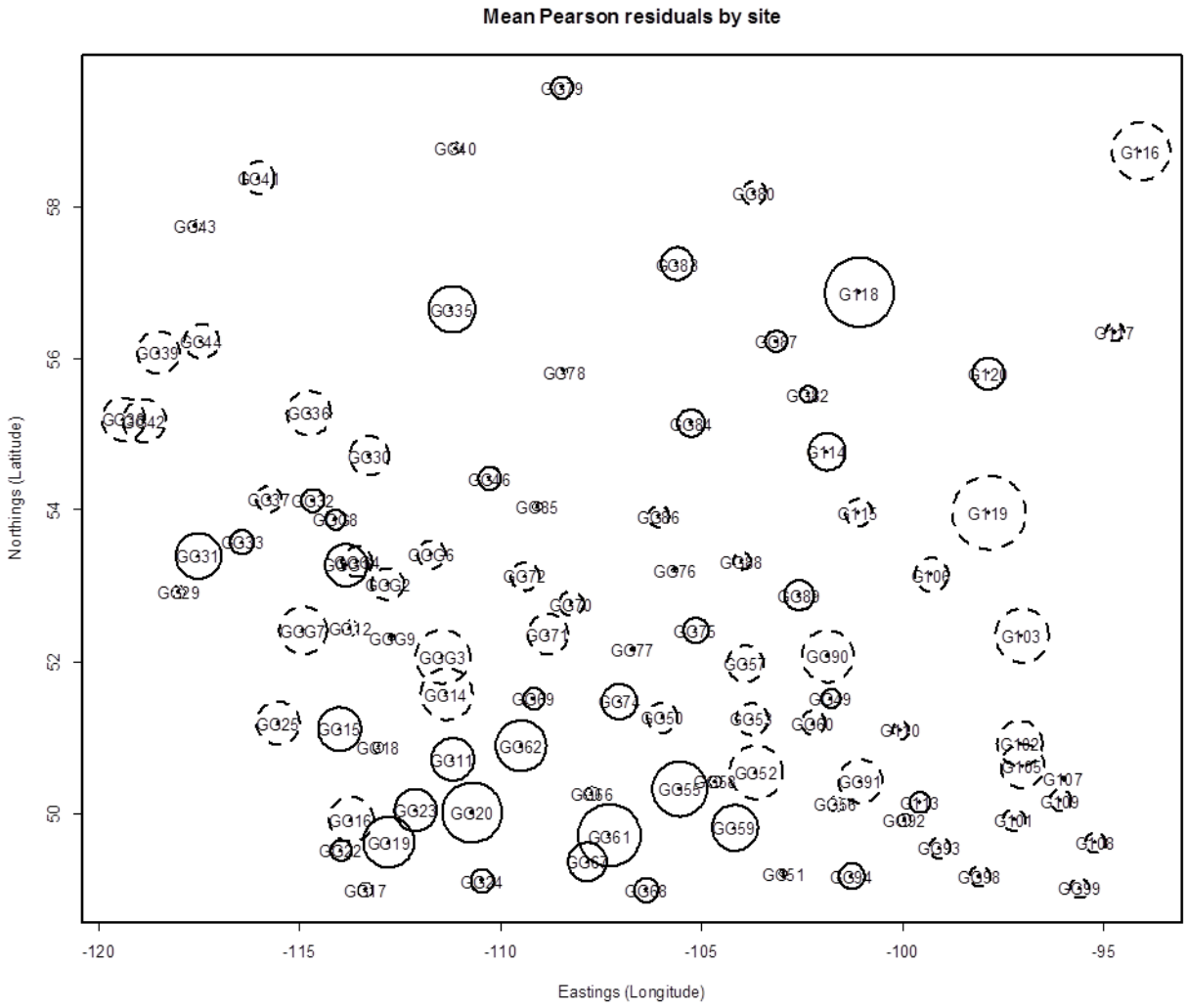

Figure B8: Bubble map showing mean residuals from the jointly fitted mean and variance model at each site in the entire study region. Circle areas are proportional to standardized mean residuals.

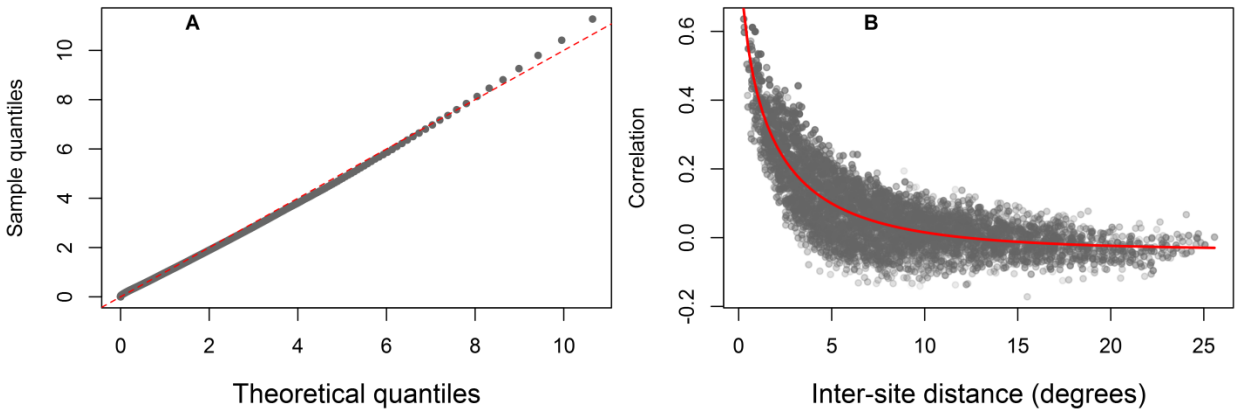

Figure B9: (A) Q-Q plot of standardized residuals from the fitted amounts model. The data values are shown as grey dots. (B) Empirical inter-site correlations (grey dots) that decay exponentially with distance; the red line is the fitted correlation model. The plot assesses the adequacy of the assumed correlation structure.

APPENDIX C: SUPPORTING MATERIAL FOR CHAPTER 2

C1.1 PCA Theory

Principal Components (PCs) are defined as the eigenvectors of the covariance matrix derived from the time series of the data field (von Storch et al., 1995; Wilks, 1995). As a data reduction technique, PCA reduces a dataset containing a large number of variables to fewer variables, but that still represent a large fraction of the variability contained in the original dataset. Thus, it attempts to simplify a complex set of inter-relationships between variables by creating one or more new variables with respect to the original ones that explain a significant portion of the original variability. According to Richman (1986); Graham (1988); and Preisendorfer and Mobley (1988), it aims at explaining the overall variance in a dataset by statistically isolating a number of components with respect to newly defined axes, with each corresponding to a variable. Thus, it defines the orthogonal coordinate system that describes the maximum variance of one data set.

For an $n * p$ matrix, $\boldsymbol{Z}$, by computing a transformation of $\boldsymbol{Z}$, say $\boldsymbol{U}$ which explains maximum variance between linear combinations of $p_1, p_1, ..., p_n$ variables such that the transformed data, $\boldsymbol{V} = \boldsymbol{ZU}$ and

$$\text{var}(\boldsymbol{V}) = \frac{\boldsymbol{V}'\boldsymbol{V}}{p\text{-}1} = \frac{(\boldsymbol{ZU})'\boldsymbol{ZU}}{p\text{-}1} = \frac{\boldsymbol{U}'\boldsymbol{Z}'\boldsymbol{ZU}}{p\text{-}1} \tag{1}$$

but $\frac{\boldsymbol{Z}'\boldsymbol{Z}}{p-1}$ is the variance-covariance matrix, $\boldsymbol{S}$, so that: $\text{var}(\boldsymbol{V}) = \boldsymbol{U}'\boldsymbol{SU}$. The Variance, $\boldsymbol{V}$, should be at maximum. Thus, a maximization problem has to be solved. By constraining $\boldsymbol{U}$ to unit length, $\boldsymbol{U}'\boldsymbol{U} = 1$ and introducing a Lagrange multiplier, the maximization problem becomes:

$$\text{var}(\boldsymbol{V}) = f(\boldsymbol{U}) = \boldsymbol{U}'\boldsymbol{SU} - \lambda\left(\boldsymbol{U}'\boldsymbol{U} - I\right) = max \text{, where}$$

$\frac{\partial f(\boldsymbol{U})}{\partial \boldsymbol{U}} = 0$ or $|\boldsymbol{S} - \lambda \boldsymbol{I}|\boldsymbol{U} = 0$ has the same structure as the characteristic equation that determines the eigen structure of a square matrix, $\boldsymbol{A}$, ($|\boldsymbol{A} - \lambda \boldsymbol{I}|E = 0$). E and λ contain the eigenvectors and eigenvalues, respectively. Thus, the transformation that maximizes the variation of $\boldsymbol{S}$ is given by its eigenvectors/principal component (PC), and the principal component with the highest eigenvalue gives the transformation with the largest variance.

Given that $\boldsymbol{U}$ was constrained to unity, the total variance explained by each eigen mode/PC is given by $\Sigma_{i=1}^{n} \lambda_i$. One of the main problems is to determine the number of PCs to maintain. There are several criteria such as the 'Scree test' (Catell, 1966) and the Kaiser criterion (Kaiser, 1958). Normally, the first PC takes into account a large part of the variability in the original dataset while other PCs can bring to light interesting phenomena hidden in $\boldsymbol{Z}$.

C1.2 Canonical Correlation Analysis (CCA)

A good description of CCA as applied in climate sciences can be found in Graham (1990) and Wilks (2011). Assume that there are two datasets $\boldsymbol{X}_{t,x}$ and $\boldsymbol{Y}_{t,y}$ where the subscripts represent time t and spaces x and y. Time t has nt time steps that must be the same for $\boldsymbol{X}$ and $\boldsymbol{Y}$. The spatial dimensions nx and ny must not be the same but the mean of each column of $\boldsymbol{X}$ and $\boldsymbol{Y}$ must be zero. By applying CCA to the normalized reconstructed fields, $\boldsymbol{X}$ and $\boldsymbol{Y}$, the purpose is to find a pair of vectors $\boldsymbol{U}$ and $\boldsymbol{V}$ that is a linear combination (LC) from $\boldsymbol{X} = [x_1, x_2, x_3, ...x_x]$ independent, and $\boldsymbol{Y}$ dependent variables, $\boldsymbol{Y} = [y_1, y_2, y_3, ...y_x]$. The CCA creates two new variables, $\boldsymbol{U} = [u_1, u_2, u_3, ..., u_n]$ and $\boldsymbol{V} = [v_1, v_2, v_3, ..., v_n]$, that are maximally correlated, where $n = \min(x, y)$. That is;

$$\boldsymbol{U} = \boldsymbol{YE} \tag{2}$$

$$\boldsymbol{V} = \boldsymbol{XQ} \tag{3}$$

and the correlation, between LCs $\boldsymbol{U}$ and $\boldsymbol{V}$

$$corr\ (\boldsymbol{U},\boldsymbol{V}) = \frac{\sum \boldsymbol{UV}}{\sqrt{\sum \boldsymbol{U}^2 \sum V^2}} \tag{4}$$

It can further be shown that if $\boldsymbol{Y}$ is the predictand variable and $\boldsymbol{E}$ is the transformation, then $\boldsymbol{U}$ is the transformed variable whose variance is derived as;

$\frac{\sum \boldsymbol{U}^2}{(n-1)} = \frac{\boldsymbol{U}'\boldsymbol{U}}{(n-1)} = \frac{(\boldsymbol{YE})'(\boldsymbol{YE})'}{(n-1)} = \frac{\boldsymbol{E}'\boldsymbol{Y}'\boldsymbol{YE}}{(n-1)}$, where $(.)'$ is the transpose.

Furthermore, if $\boldsymbol{Y}'\boldsymbol{Y}/(n_t - 1)$ is the covariance matrix ($\boldsymbol{S}_{yy}$) of $\boldsymbol{Y}$, then,

$$\sum \boldsymbol{U}^2 = \boldsymbol{E}'\boldsymbol{S}_{yy}\boldsymbol{E} \tag{5}$$

In the case of $\boldsymbol{X}$ and the cross-covariance between $\boldsymbol{X}$ and $\boldsymbol{Y}$:

$$\sum \boldsymbol{V}^2 = \boldsymbol{Q}'\boldsymbol{S}_{xx}\boldsymbol{Q} \tag{6}$$

$$\sum \boldsymbol{UV} = \boldsymbol{E}'\boldsymbol{S}_{yx}\boldsymbol{Q} \tag{7}$$

Note that $\boldsymbol{S}_{yy} = \boldsymbol{Y}'\boldsymbol{Y}/(nt-1); \boldsymbol{S}_{xx} = \boldsymbol{Z}'\boldsymbol{Z}/(nt-1);$ and $\boldsymbol{S}_{yx} = \boldsymbol{Y}'\boldsymbol{X}/(nt-1)$. By substituting (5), (6) and (7) in (4) gives:

$$corr(\boldsymbol{U},\boldsymbol{V}) = \frac{\boldsymbol{E}'\boldsymbol{S}_{yx}\boldsymbol{Q}}{\sqrt{\boldsymbol{E}'\boldsymbol{S}_{yy}\boldsymbol{E}\ *\boldsymbol{Q}'\boldsymbol{S}_{xx}\boldsymbol{Q}}} \tag{8}$$

The problem is to maximize (8) subject to the constraint of total variance being equal to 1, i.e.,

$\boldsymbol{E}'\boldsymbol{S}_{yy}\boldsymbol{E}=1$ and $\boldsymbol{Q}'\boldsymbol{S}_{xx}\boldsymbol{Q}=1$. We use the method of Lagrange multipliers to solve the problem.

$\boldsymbol{F}\left(\boldsymbol{E},\boldsymbol{Q}\right) = \boldsymbol{E}'\boldsymbol{S}_{yx}\boldsymbol{Q}$ is the function to be maximized, $\boldsymbol{E}'\boldsymbol{S}_{yy}\boldsymbol{E}=1$ and $\boldsymbol{Q}'\boldsymbol{S}_{xx}\boldsymbol{Q}=1$ are the constraints, $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are Lagrange multipliers. Thus, we can write

$\boldsymbol{F}\left(\boldsymbol{E},\ \boldsymbol{Q}\right) = \boldsymbol{E}'\boldsymbol{S}_{yx}\boldsymbol{Q} - \mu\ \left(\boldsymbol{Q}'\ \boldsymbol{S}_{xx}\boldsymbol{Q} - 1\right) - \lambda\left(\boldsymbol{E}'\boldsymbol{S}_{yy}\boldsymbol{E} - 1\right)$ and setting the derivatives $\partial\boldsymbol{F}/\partial\boldsymbol{E}'=0;\ \partial\boldsymbol{F}/\partial\boldsymbol{Q}=0$ it is possible to compute the maximum values of $\boldsymbol{F}$ **:**

$$\partial F/\partial E'=\boldsymbol{S}_{yx}\boldsymbol{Q}-\ \boldsymbol{\lambda}S_{yy}R=0 \tag{9}$$

$$\partial F/\partial Q=\boldsymbol{E}'\boldsymbol{S}_{yx}-\boldsymbol{\mu}Q'S_{xx}=0 \tag{10}$$

Let's multiply (9) by $\boldsymbol{E}'$ and post-multiply (10) by $\boldsymbol{Q}$ to get $\boldsymbol{E}'\ \boldsymbol{S}_{yx}\boldsymbol{Q}=\boldsymbol{E}'\lambda\boldsymbol{S}_{yy}\boldsymbol{E}=\mu\boldsymbol{Q}'\boldsymbol{S}_{xx}\boldsymbol{Q}$, but $\boldsymbol{E}'\boldsymbol{S}_{yy}\boldsymbol{E}=\boldsymbol{Q}'\boldsymbol{S}_{xx}\boldsymbol{Q}=1$, so

$$\boldsymbol{E}'\boldsymbol{S}_{yx}\boldsymbol{Q}=\lambda=\mu\ = \text{maximum correlation}\ (\boldsymbol{\rho}) \tag{11}$$

By re-substituting (11) in (9) and (10), we get

$$\boldsymbol{S}_{yx}\boldsymbol{Q}=\mu\,\boldsymbol{S}_{yy}\boldsymbol{E} \tag{12}$$

$$\boldsymbol{S}_{xy}\boldsymbol{E}=\mu\boldsymbol{S}_{xx}\boldsymbol{Q} \tag{13}$$

As $\boldsymbol{Q}'\boldsymbol{S}_{xx}=\boldsymbol{S}_{xx}\boldsymbol{Q}$, we can use (13) to get

$$\boldsymbol{Q}=1/\mu\ \boldsymbol{S}_{xx}-1\boldsymbol{S}_{xy}\boldsymbol{E} \tag{14}$$

A combination of (14) and (12) results in

$\boldsymbol{S}_{yx}\ (1/\mu\ \boldsymbol{S}_{xx}-1\boldsymbol{S}_{xy}\boldsymbol{E})=\mu\,\boldsymbol{S}_{yy}\boldsymbol{E}$. We pre-multiply this equation by $\mu\boldsymbol{S}_{yy}^{-1}$ and by rearranging the terms, we obtain $(\boldsymbol{S}_{yy}-1\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}-1\boldsymbol{S}_{xy}-\mu^2\boldsymbol{I})\boldsymbol{E}=0$ where μ^2 represents the eigenvalues of $(\boldsymbol{S}_{yy}-1\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}-1\boldsymbol{S}_{xy})$, $\boldsymbol{E}$ represents the eigenvectors, and $\boldsymbol{I}$ is the identity matrix. It is logical from here that equation (13) gives the solution for $\boldsymbol{Q}$ and $\boldsymbol{E}$, so it is possible to obtain the canonical time series (vectors) of these canonical spatial patterns $\boldsymbol{E}$ and $\boldsymbol{Q}$ by

$$\boldsymbol{U}_{nt,nm}=\boldsymbol{Y}_{nt,ny}\,\boldsymbol{E}_{ny,nm} \tag{15}$$

and

$$\boldsymbol{V}_{nt,nm}=\boldsymbol{X}_{nt,nx}\boldsymbol{Q}_{nx,nm} \tag{16}$$

The transformed variables, $\boldsymbol{V}$ and $\boldsymbol{U}$ are paired so that $\boldsymbol{V}_1$ and $\boldsymbol{U}_1$ are correlated with the canonical correlation coefficient ρ_1, and $\boldsymbol{V}_2$ and $\boldsymbol{U}_2$ are correlated with ρ_2 and so on.

Table C1: Values of the heterogeneity measures H_k for each of the five regions (after defuzzification) for the case of winter and summer seasonal precipitation totals

Region	Winter			Summer		
	H_1	H_2	H_3	H_1	H_2	H_3
A	0.69	0.85	-0.57	0.84	0.90	0.34
B	0.91	0.40	0.80	0.97	0.68	0.67
C	0.96	-0.26	0.52	0.89	0.88	0.60
D	0.53	0.50	0.61	-0.34	-0.14	1.32
E	0.51	-0.74	0.05	0.29	0.63	-0.41

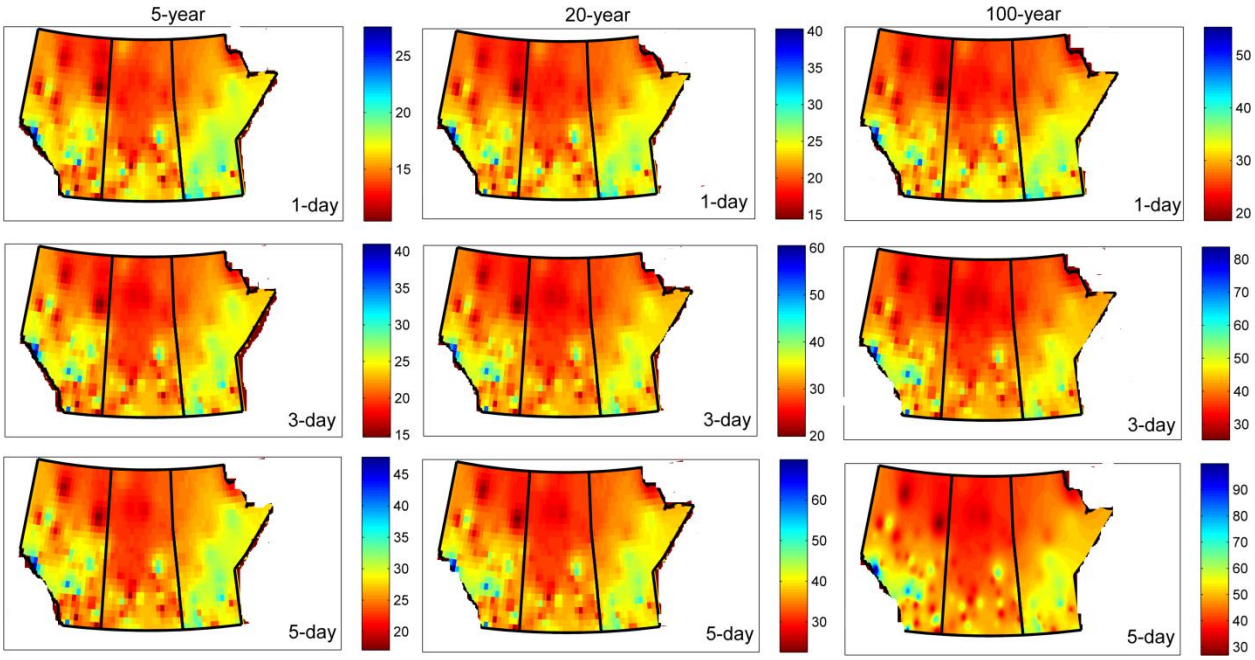

Figure C1: Spatial distribution of 5- (column 1), 20- (column 2) and 100-year (column 3) return values of 1-, 3-, and 5-day winter season precipitation extremes (in mm)

APPENDIX D: PERMISSIONS FOR USE OF PUBLISHED MANUSCRIPTS

SPRINGER LICENSE
TERMS AND CONDITIONS

Sep 03, 2015

This is a License Agreement between ZILEFAC E ASONG ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3701560840244
License date	Sep 03, 2015
Licensed content publisher	Springer
Licensed content publication	Stochastic Environmental Research and Risk Assessment
Licensed content title	Regionalization of precipitation characteristics in the Canadian Prairie Provinces using large-scale atmospheric covariates and geophysical attributes
Licensed content author	Z. E. Asong
Licensed content date	Jan 1, 2014
Volume number	29
Issue number	3
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	1
Author of this Springer article	Yes and you are the sole author of the new work
Order reference number	None
Title of your thesis / dissertation	MULTIVARIATE MULTISITE STATISTICAL DOWNSCALING OF ATMOSPHERE-OCEAN GENERAL CIRCULATION MODEL OUTPUTS OVER THE CANADIAN PRAIRIE PROVINCES
Expected completion date	Oct 2015
Estimated size(pages)	156
Total	0.00 USD
Terms and Conditions	

Introduction
The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

Limited License
With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry.

Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: http://www.sherpa.ac.uk/romeo/). For any other electronic use, please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com).

The material can only be used for the purpose of defending your thesis limited to university-use only. If the thesis is going to be published, permission needs to be re-obtained (selecting "book/textbook" as the type of use).

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, subject to a courtesy information to the author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well).

Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

Altering/Modifying Material: Not Permitted
You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com)

Reservation of Rights
Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

Copyright Notice:Disclaimer
You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media"

Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity
You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms
Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration).***OR:***

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other terms and conditions:

v1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.